Reason to Believe

Chenwei Shi

Reason to Believe

ILLC Dissertation Series DS-2018-10

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Science Park 107
1098 XG Amsterdam
phone: +31-20-525 6051
e-mail: illc@uva.nl
homepage: http://www.illc.uva.nl/

This work was financially supported and hosted by the Tsinghua-UvA Joint Research Center for Logic.

Cover design by Chenwei Shi.
Printed and bound by Ipskamp Printing.

ISBN: 978–94–028–1032–5

Reason to Believe

Academisch Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde
commissie, in het openbaar te verdedigen in de Aula der Universiteit
op woensdag 13 juni 2018, te 13.00 uur

door

Chenwei Shi

geboren te Zhejiang, China

Promotiecommisie

Promotor:	Prof. dr. S.J.L. Smets	Universiteit van Amsterdam
Promotor:	Prof. dr. F. Liu	Universiteit van Amsterdam
Overige leden:	Prof. dr. J.F.A.K. van Benthem	Universiteit van Amsterdam
	Prof. dr. F. Berto	Universiteit van Amsterdam
	Dr. P. Egré	Institut Jean-Nicod
	Prof. dr. D.J.N. van Eijck	Universiteit van Amsterdam
	Prof. dr. D. Grossi	Rijksuniversiteit Groningen
	Prof. dr. B. Liao	Zhejiang University
	Prof. dr. F.J.M.M. Veltman	Universiteit van Amsterdam

Faculteit der Geesteswetenschappen

To my parents, Fang Shi and Weidi Yang.

Contents

Acknowledgments

It's been a long journey. Without the support and help from many kind people, it is never possible to even start, not to mention reaching the current point.

First of all, I want to thank my official or unofficial supervisors, Sonja Smets, Fenrong Liu, Johan van Benthem, and Fernando R. Velázquez Quesada.

Without Johan, I couldn't have started my PhD here at ILLC. For anyone, it would be a venture to accept a student like me who had not been exposed to enough education on logic. I always felt lack of enough training in logic after I came to ILLC, especially compared to students in the Master of Logic Programme. However, Johan always patiently listened to my naive research ideas and tried to guide me to the right directions. Such guidance runs through the whole journey and leads this thesis to its current shape.

Without Sonja, I couldn't have persisted on this journey. Sonja is the first person here who I feel I can always appeal to for help when something goes wrong. The consequence for Sonja, I guess, is more experience with all kinds of unexpected and absurd troubles. It is in the first joint project with Sonja that I learn how to conduct research properly, discussing and clarifying a research idea, putting the idea into words, and presenting it in an accessible way.

Without Fenrong, I couldn't have even known what logic is. In her course, I ran into formal logic and was amazed by its rigour and clarity. It is Fenrong that firstly teaches me to appreciate the beauty of logic. It is also Fenrong that points out the connection between formal argumentation theory, belief revision and social network to me. Fenrong is always so encouraging even when I almost lost patience with myself.

I also feel gratitude to Fernando R. Velázquez Quesada. It is after he came to ILLC and with his help that I started feeling that there was still hope of graduating on time. It is such a pleasure to work with Fer. He is willing to listen to my vague ideas and read my sloppy proofs. Moreover, he is good at making things clear and rigorous. I learn a lot from Fer: the writing and presentation skills, his insight on different research topics, and his attitude towards research.

Fer is like a brother who I can share not only my happiness but also my worries and complaints.

Second, I would like to thank the members of my thesis committee: Johan van Benthem, Franz Berto, Paul Egré, Jan van Eijck, Davide Grossi, Beishui Liao and Frank Veltman, who kindly agreed to commit their time to reading and assessing this thesis.

Third, I want to express my gratitude towards Alexandru Baltag and Olivier Roy, with whom I had the honour to work. After being Alexandru's colleague for four years, I have already been used to his genius. However, his passion for alcohol is still a mystery to me, especially when it coexists with his passion for logic. Olivier is also an ardent logician. I appreciate the chance of visiting Bayreuth offered by Olivier. The discussions there with Olivier and his group — Albert Anglberger, Huimin Dong, Marcel Heidemann and Dominik Klein — are my memorable experience.

Fourth, I am very grateful to ILLC for creating the perfect environment for research. I often heard in China that the country planned to construct "world-class universities", of which I totally had no idea. ILLC helps turn this abstract concept into my concrete experience, for which, the members of ILLC management team and ILLC office deserve my sincere gratefulness (in particular, I want to thank Jenny for translating the title of my thesis into Dutch together with Sonja). My gratitude also extends to the members of the PhD council, Samira Abnar, Joost Bastings, Dieuwke Hupkes, Arnold Kochari and Marco Del Tredici, who taught me how the students from an institute could contribute to its success not only in a passive way.

To be a member of Amsterdam Dynamics Group is definitely a very dynamic experience, concerning the variety of its research topics and its members' talents. The people in the following incomplete list witness to such a variety: Soroush Rafiee Rad, Ilaria Canavotto, Giovanni Ciná, Paolo Galeazzi, Malvin Gattinger, Aybüke Özgün, Ana Lucia Vargas Sandoval, Kaibo Xie, Roberto Ciuni, Zoé Christoff, Nina Gierasimczuk, Shunan He, Chanjuan Liu, Joshua Sack, Shengyang Zhong, Bryan Renne, Rasmus K. Rendsvig, and so on. They constitute a great proportion of my life at ILLC. I learn a lot from them, for example, the topological semantics for evidence and belief from Aybüke, game theory from Paolo, stit from Ilaria, and how to be an organised person from Malvin among so many other things.

I also feel lucky to share offices during the four years with Samira Abnar, Zoé Christoff, Paolo Galeazzi, Ilaria Canavotto, Thom van Gessel, Gianluca Grilletti, Tom Schoonen, Anthia Solaki and Kaibo Xie. Samira, thank you for having tea breaks with me and sharing those Iranian sweets; Paolo, thank you for the house and music; Thom, thank you for answering my questions about politics in the Netherlands; Tom, thank you for translating the abstract for my thesis; Kaibo, thanks for your humour. Thank you all for your kindness and tolerance.

Fifth, I want to mention the Chinese community who were or are working and studying at or around ILLC during my stay in Amsterdam: Weiwei Chen, Yibin Dai, Xiaoxuan Fu, Tao Gu, Lifeng Han, Shunan He, Haoqian Jiang, Fengkui Ju, Ko-Hung Kuan, Kai Li, Yanjun Li, Zhenhao Li, Chanjuan Liu, Yuning Liu, Ciyang Qing, Xin Sun, Chong Wang, Shuai Wang, Tong Wang, Yanjin Wang, Kaibo Xie, Yujie Xing, Fan Yang, Yuan-ho Yao, Junhua Yu, Fangzhou Zhai, Bonan Zhao, Zhiguang Zhao, Zhuoye Zhao, Tingxiang Zou. Thank you all for the time we spent together, which makes my life here colourful.

Also, my thanks go to John Gu for his wonderful stories and Michiel den Haan for his warm help during my exchange at ILLC and at the very beginning of my PhD study. I also owe a lot of thanks to my flatmates: Klim Mikhailov, Shahid Sumon, Haoqian Jiang, Persofoni Bernadou, and Malvin Gattinger. They make those empty and cold flats home to me.

Special thanks go to Chenchen, who has been providing a home for my heart, without which I could have fallen apart.

At last, I dedicate this thesis to my parents, Fang Shi and Weidi Yang. As parents, they have always questioned and worried about my choices and opinions. However, they never stop me from doing what I want. Instead, they support me conditional on nothing. Thank you for your love!

Despite these people's huge contribution to my PhD study and this thesis, without any doubt, I should bear all the responsibility for mistakes and omissions in this thesis.

Chenwei Shi
April, 2018.

Chapter 1

Stratification of the Doxastic Earth

1.1 Introduction

This dissertation studies belief and its relations to evidence and reasoning from a logician's perspective. We deal with not only the notion of a single agent's belief but also the notion of a group's collective belief.

1.1.1 Belief: a journey from chaos to order

Belief plays a central role in our daily life: what we believe can influence our decisions and behaviours to a great extent. For example, if people are confident about the development of the economy and thus the increase of their future income, they tend to spend more money now. The concept of belief, therefore, receives much attention in a number of scientific investigations. Decision theory (Savage, 1954; Jeffrey, 1965), by and large, studies how belief, together with other various attitudes, coherently contributes to an agent's choices. Moreover, in epistemology (Gettier, 1963; Goldman, 1967; Lehrer and Paxson, 1969; Dretske, 1971; Nozick, 1981), which mainly studies the notion of knowledge, most theories rely on the notion of belief.

In this dissertation, we zoom in on the notion of belief itself rather than its role in different theories. While borrowing insights and techniques from different fields, we study various aspects of belief mainly from a logician's perspective.

For logicians who follow the approach developed by Hintikka (1962) to the study of belief, consistency seems to be the most undoubted property of an agent's beliefs and thus should be encoded into its semantical definition. The following example illustrates the intuition behind the requirement of consistency.

> A rabbi is holding court in his village. Schmuel stands up and pleads his case, saying, "Rabbi, Itzak runs his sheep across my land every day and it is ruining my crops. It's my land. It's not fair."

> The rabbi says, "You're right!"
>
> But then Itzak stands up and says, "But Rabbi, going across his land is the only way my sheep can drink water from the pond. Without it, they'll die. For centuries, every shepherd has had the right of way on the land surrounding the pond, so I should too."
>
> And the rabbi says, "You're right!"
>
> The cleaning lady, who has overheard all this, says to the rabbi, "But, Rabbi, they can't both be right!"
>
> And the rabbi replies, "You're right!" [1]

Lack of consistency in one's belief can lead to absurdity and sometimes even catastrophic consequences. So it seems natural to request consistency of beliefs, not only for logicians.[2] [3] However, consistency is not a goal that is within easy reach for human beings, and it requires effort to maintain consistency of one's beliefs. Consistent beliefs do not come out of blue. Furthermore, consistent beliefs do not come out of nothing. What an agent believes depend on what (s)he encounters, perceives and receives. Information floods towards everyone in this world, it is raw and chaotic. As in the previous story, different pieces of information coming from the two sources (Schemuel and Itzak) contradicts each other. Moreover, sometimes, information coming from one source can contradict itself.

> In the state of Chu lived a man who sold shields and spears. "My shields are strong," he boasted, "that nothing can pierce them. My spears are so sharp that there is nothing they cannot pierce."
>
> "What if one of your spears strikes one of your shields?" someone asked him.
>
> The man had no answer to that.
>
> (*Han Fei Zi*)[4]

The merchant's words are just unacceptable because of its self-contradiction. Without any processing, consistency is hard to find in information received by the agent. The agent needs to process received information to achieve consistency in her beliefs. Among those pieces of information which are disqualified as belief

[1] Cathcart and Klein (2007, p.28)

[2] A celebrated argument for consistency of belief is the Dutch Book Argument which traces to Ramsey (1926), although it is mainly about degrees of belief and probabilistic consistency.

[3] Concerning the concept "consistency" itself, although the Law of Non-Contradiction introduced by Aristotle is conceived as an orthodox principle, there is still room of contending for its opposite. For such an admiring endeavour, see Berto (2007).

[4] Xianyi and Naidie (2016)

must be the self-contradictory ones. We will call a body of propositions which does not include any self-contradictory proposition *1-consistent*.

Similarly, two contradictory propositions should not coexist as an agent's beliefs, since the contradiction may come from their intersection. When the rabbi says "You're right" to both Schmel and Itzak, there is nothing wrong for the rabbi to accept their words as evidence at the same time so that he takes both of them into account; the absurdity comes from accepting both of them as his beliefs. We will call a body of propositions which does not include any pair of contradictory propositions *2-consistent* or *mutually consistent*.

In doxastic logic, 2-consistency is still not consistent enough. In fact, doxastic logic requires that the agent's belief should be ∞-consistent or fully consistent, that is, any finite subset of what the agent believes should not lead to contradiction.[5]

From 1-consistency to ∞-consistency, it is reasonable to think of different propositional attitudes besides belief with different levels of consistency. For example, one can think of other kinds of attitudes such as "taking or considering something as evidence". Towards different propositions, the agent may adopt different attitudes with different levels of consistency, depending on how committed she is. After all, there is no need to take every proposition equally serious. We call all these propositional attitudes "*doxastic attitudes*".

The agent's doxastic attitudes, like the earth, can be stratified into different layers according to different levels of consistency (Figure 1.1). The core of the earth thus contains those attitudes which require ∞-consistency; and the attitudes in the surface layer require 1-consistency. Information, like rain, falls to the earth. Once touching the ground, it permeates through the earth. The deeper it penetrates, the less of it remains.

In this dissertation, we will dig into the doxastic earth. We start from the surface layer, where the agent's evidence lives. An agent can accept two mutually contradictory pieces of information as evidence but would not accept any self-contradictory information as evidence. So an agent's evidence should be 1-consistent. Since the agent's beliefs should be more consistent than 1-consistency, we have to dig deeper. How should the agent's beliefs be based on her evidence so that her beliefs are consistent enough? The first half of this dissertation is devoted to exploring how *a single agent* achieves this by different forms of reasoning.

1.1.2 Reasoning: bottom up

In doxastic logic (Hintikka, 1962), full consistency is the starting point of the whole enterprise of a logical study of belief. In such a context, reasoning is usually taken as a way of deriving new information from what an agent consistently believes, which is embodied in the logical properties of belief. For example, if the

[5] A more formal definition of these notions of consistency can be found in Section 1.2.

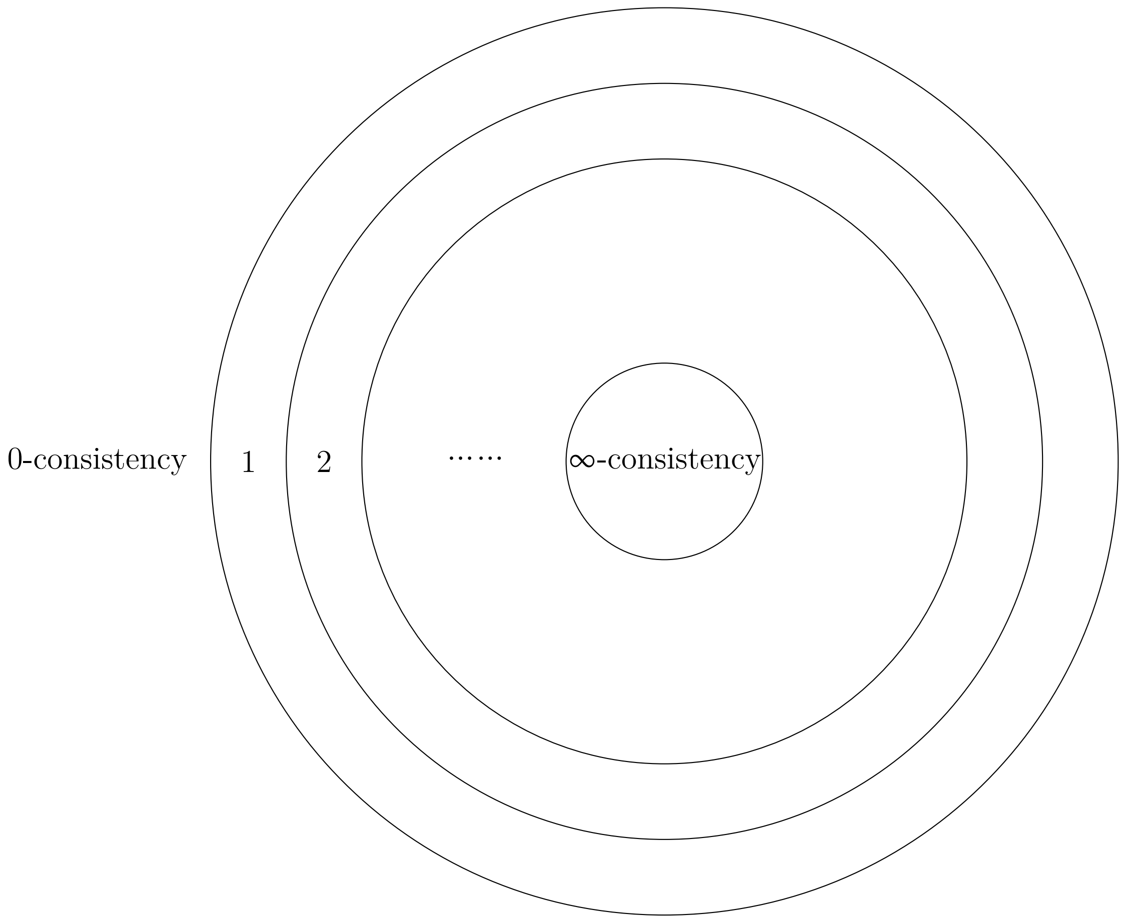

Figure 1.1: Doxastic earth with layers of consistency

agent believes that A and believes that A implies C, then she would believe that C. So reasoning brings out what is implicit in the agent's beliefs.

When belief is assumed to be at the core of the doxastic earth, the function of reasoning is also restricted within the core. However, in this dissertation, our journey does not start from the core, but rather from the surface layer of the doxastic earth. We do not take a fully consistent notion of belief as given but as a goal to be reached. So it does not suffice anymore only to maintain the consistency within the top level of the hierarchy of consistency. It is equally important to resolve the inconsistency in the agent's evidence.

The question of how the agent reasons to resolve inconsistency is closely related to the question of how consistent the agent's belief should be. Is 2-consistency consistent enough for belief? If not, does it have to be ∞-consistent? Although full consistency seems to be the most undoubted requirement in the modal semantics of belief in doxastic logic, it is the deepest core in the doxastic earth and therefore the hardest level to reach in the hierarchy of consistency.

Concerning the role of belief in decision making in practice, there are reasons to compromise some consistency. For example, the complexity of achieving and maintaining consistency on a full scale might be overly high; the inconsistency is sometimes not that relevant to the task at hand and so on. As these practical factors vary, different levels of consistency may be required.

From a purely logical perspective, full-consistency of belief is not an unquestionable property either. Since the process of achieving consistency is a process of sifting out more and more information which causes inconsistency, the higher the consistency-level is required, the more information should be thrown away. In the doxastic earth, the deeper the rain permeates underground, the less of it remains. Hence it is possible that too few propositions can be believed if the agent's beliefs have to be fully consistent. There is a tension between believing more and believing more consistently. It is crucial that the agent can strike a balance between these two points. The question is whether we have to sacrifice some consistency to avoid leaving too few propositions to be believed. The first half of this dissertation attempts to answer this question in a formal framework.

1.1.3 Group belief: towards compromise or consensus via social interaction

In the case of a single agent, the agent's belief directs her actions. What then directs the actions of a group, for example, the investment choices of an investment bank? Group belief? But what is group belief? What does it mean to say "the board of the bank believes...", "the government believes..." and so on?

Instead of pursuing these questions purely philosophically (Gilbert, 1987; List and Pettit, 2011; Bratman, 2014), we approach the problem from a more practical perspective. In this dissertation, we adopt the working hypothesis that no matter what group belief can be, it should be based on a consensus or at least a compromise between the group members. So the central question we try to tackle in the second part of this dissertation is how the members of a group with different and even incompatible beliefs can reach a consensus or compromise.

The question bears an apparent similarity to the question of how a single agent resolves contradictions between different pieces of evidence to achieve more consistent beliefs. However, the differences between the single and multi-agent setting are also evident. Firstly, a group cannot reason in the same way as a single agent. It cannot be expected that a group can settle incompatibility between different agents' beliefs as spontaneously as a single agent resolves the contradiction between different pieces of evidence. More effort is needed for a group to coordinate its members than for a single agent to play around with her evidence. So some external method should be used to achieve the same effect for a group of agents as reasoning for a single agent, for example, voting. Secondly, as a result of the first difference, those methods should be carried out issue by issue, rather than holistically.

In this dissertation, rather than focusing on methods such as voting, we study two other ways of reaching consensus or compromise within a group of agents, based on which our notions of group belief will be defined. Both of these two ways emphasise the role of social interactions between group members in the formation of the group's beliefs.

The first way to define group belief is through argumentation, by which we mean not only the dynamic process of a debate among group members but also the static structure of attack relation between arguments abstracted from the debate. Compared with voting, argumentation is a more discursive method, which can be taken as a form of deliberation, as advocated in the literature on deliberative democracy (Habermas, 1996; Dryzek, 2000; Gutmann and Thompson, 2004). Distinctively, we do not pursue how group members influence each other in argumentation and whether argumentation as a deliberative process would help to produce unanimity among group members. What we will study about argumentation is merely a way of deciding the winning side, given structural information on the attack relation between group members. As for the mutual influence through social interaction between group members and its role in group belief, the issue will be addressed in the second way of defining group belief.

Instead of working with the attack relation in a debate, the second way of defining group belief considers how much one group member trusts the other members and how this relationship between group members would influence their beliefs in the long run. To be more specific, the second way of reaching consensus is through a process of opinion diffusion which is triggered by the mutual influence between group members. However, as Keynes famously put it, "in the long run we are all dead" (Keynes, 1923, Chapter 3, p.80). It is not at all practical to make a group's current belief dependent on its far-future consensus. So instead of the realisation of the far-future consensus, we base group belief on the *tendency* of a far-future consensus. To know this tendency, there is no need to "wait and see". Tendency depends purely on the structure of the current trust network of the group. Although knowing how the group members' belief would change may not suffice for deciding the current action of the group, there is no doubt that it is helpful as a guide. In this sense, the tendency is suitable to be taken as the basis of group belief.

For group belief, how consistent it needs to be is also an important question. We will see that the tension between believing more and believing more consistently recurs for group belief.

I conclude this introduction with an overview of each chapter.

Chapter two starts with a review of the notion of evidence-based belief defined in the evidence model (van Benthem and Pacuit, 2011) and the notion of justified belief defined in the topological semantics for evidence (Baltag, Bezhanishvili, Özgün and Smets, 2016*a*). By combining the topological semantics for evidence and formal argumentation theory (Dung, 1995), we propose a new setting to model a notion of *argument*-based belief called "*grounded belief*". The comparison between justified belief and grounded belief then reveals the tension we have just introduced. Grounded belief sacrifices consistency for more content, while justified belief sacrifices content for more consistency.

In **Chapter three**, we try to relieve the tension between believing more and believing more consistently by looking for a notion of belief which would not sacrifice consistency for more content and would not be as stringent as justified belief. This goal is accomplished by coordinating argumentational reasoning with *default reasoning*, which is represented by the specification order generated from a topological space. The resulting notion of belief is called "full-support belief". We do not claim that this is the best notion of belief, but rather demonstrate that there is a possibility of striking a balance and that the apparent tension is not a real dilemma.

Chapter four turns to group belief based on argumentation among the group members. Formal argumentation theory is integrated into multi-agent doxastic logic to model the argumentation and the notion of group belief based on it. The way of applying formal argumentation theory to a notion of group belief called "argumentation-based group belief" is quite different from what is discussed in chapter two for grounded belief. But the tension that exists for grounded belief also exists for argumentation-based group belief.

Chapter five approaches group belief from a different angle. *Markov chain theory* serves as the technical basis for the work in this chapter, which is used to model the tendency of the change of the group members' beliefs. A notion called "potential group belief" is thus defined based on the tendency and we study the logic of this notion of group belief. At the end of this chapter, we elaborate on the connection between group belief and judgement aggregation.

At last, we conclude the whole dissertation and offer some open questions for future work in **Chapter six**.

1.1.4 Sources of each chapter

Chapter two is based on:

> Chenwei Shi, Sonja Smets and Fernando R. Velàzquez-Quesada. 2017a. "Argument-based belief in topological structures." Electronic Proceedings in Theoretical Computer Science 251: pp. 489–503. In Proceedings TARK 2017, arXiv: 1707.08762.
>
> Chenwei Shi, Sonja Smets and Fernando R. Velàzquez-Quesada. 2017d. Logic of argument and belief. Institute for Logic, Language and Computation, University of Amsterdam.

Chapter three is based on the following manuscript in preparation:

> Chenwei Shi, Sonja Smets and Fernando R. Velàzquez-Quesada. 2017c. Default reasoning in topological semantics. Institute for Logic, Language and Computation, University of Amsterdam.

Chapter four is based on:

> Chenwei Shi, Sonja Smets and Fernando R. Velàzquez-Quesada. 2017b. "Beliefs supported by binary arguments." To appear in Journal of Applied Non-classical Logics.

For the above four papers, Chenwei Shi initiated the project, the results in the papers were discussed together with Sonja Smets and Fernando R. Velàzquez-Quesada, and all three authors contributed to writing, revising, and finalizing the papers. Chapter five is based on the following manuscript in preparation:

> Alexandru Baltag, Fenrong Liu, Chenwei Shi and Sonja Smets. 2017. Indeterministic DeGroot model and potential group belief. Institute for Logic, Language and Computation, University of Amsterdam.

For the above paper, Alexandru Baltag, Fenrong Liu, and Sonja Smets initiated the project, Chenwei Shi contributed to the writing of the paper and the results in the paper were discussed with Alexandru Baltag, Fenrong Liu, and Sonja Smets.

In addition, discussions in the following papers contribute to the development of the ideas in this dissertation to a great extent:

> Chenwei Shi and Olivier Roy. 2017. Reason to believe. In *Proceedings of Logic, Rationality, and Interaction: 6th International Workshop, LORI 2017*, ed. Alexandru Baltag, Jeremy Seligman and Tomoyuki Yamada. Berlin, Heidelberg: Springer pp. 676-680.

Chenwei Shi. 2016. Multi-agent epistemic argumentation logic. In Proceedings of the ESSLLI 2016 Student Session, pp. 111-122.
URL: http://esslli2016.unibz.it/wp-content/uploads/2016/09/esslli-stus-2016-proceedings.pdf

1.2 Preliminaries on Doxastic Logic

In this section, we introduce (multi-agent) doxastic logic and some notation that we will use in the chapters to follow.

We start with the relational semantics or the Kripke semantics for belief.

1.2.1. DEFINITION (Kripke frame). A Kripke frame $\mathcal{F}$ is a structure (W, R) where

- W is a non-empty set of possible worlds;
- $R \subseteq W \times W$ is a binary relation on W;

In this dissertation, whenever given a set of possible worlds, we may take each of its subsets as the semantical meaning of a proposition, called "S-proposition". For example, in a Kripke frame, we can take $P \subseteq W$ as an S-proposition. So we can also define the propositional connectives upon S-propositions as follows:

$$\neg P := W \setminus P \qquad\qquad P \wedge Q := P \cap Q$$
$$P \vee Q := P \cup Q \qquad\qquad P \to Q := \neg P \vee Q$$

Sometimes we will use notation $\overline{P}$ for $\neg P$.

The binary relation R is usually called an "accessibility relation". In the case where R represents "doxastic accessibility", Rwv is interpreted as world v is considered doxastically possible for the agent in world w. The agent's belief of a proposition, say $P \subseteq W$, is then defined as

$$BP = \{w \in W \mid \text{for all } v \in W \text{ such that } Rwv, v \in P\} \ .$$

Thus, an agent believes proposition $P \subseteq W$ if and only if all the worlds she considers doxastically possible belong to P.

Without any conditions on the doxastic accessibility relation, the logical behaviour of the belief operator B can be unbridled. For example, believing a contradiction is allowed if R is not restricted by any conditions, which may not be very desirable. So usually the accessibility relation R is required to be serial (for all $w \in W$ there is $v \in W$ such that Rwv) so that $B\varnothing = \varnothing$ always holds. Also, transitivity (for all $w, v, u \in W$, if Rwv and Rvu then Rwu) and Euclideanity (for all $w, v, u \in W$, if Rwv and Rwu then Rvu) are also deemed reasonable properties of the doxastic accessible relation, because they ensure that the agent is fully introspective about her belief: for every $P \subseteq W$

$$BP \to BBP \qquad\qquad \neg BP \to B\neg BP$$

The basic modal language is defined as:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid B\varphi$$

where p belongs to a given set of atomic propositions At. Formulas in the language can be evaluated in a Kripke model $\mathcal{M} = (\mathcal{F}, V)$, which is formed by adding a valuation function $V : \mathsf{At} \to 2^W$ to a Kripke frame $\mathcal{F}$ (2^W denotes the power set of W).

$\mathcal{M}, w \vDash p$	iff	$w \in V(p)$;
$\mathcal{M}, w \vDash \neg\varphi$	iff	$\mathcal{M}, w \nvDash \varphi$;
$\mathcal{M}, w \vDash \varphi \wedge \psi$	iff	$\mathcal{M}, w \vDash \varphi$ and $\mathcal{M}, w \vDash \psi$;
$\mathcal{M}, w \vDash B\varphi$	iff	for all $v \in W$ such that Rwv, $\mathcal{M}, w \vDash \varphi$.

In this dissertation, we will write $[\![\varphi]\!]^{\mathcal{L}}_{\mathcal{M}}$ to denote the possible worlds in a model $\mathcal{M}$ which satisfy the truth condition of $\varphi \in \mathcal{L}$. In general we will use $\mathcal{M}$ as the notation for a model in which we evaluate the sentences of a specific language. In the chapters to follow, we will work with different types of models and different types of languages. Since which type of model or language we consider should be clear in each chapter, we will usually leave out the indexes in $[\![\varphi]\!]^{\mathcal{L}}_{\mathcal{M}}$.

Satisfaction of a formula φ in a model $\mathcal{M}$ means that there is a world w in the model $\mathcal{M}$ such that $\mathcal{M}, w \vDash \varphi$; and validity of a formula φ in a class of models, denoted by $\vDash \varphi$, means that for every model $\mathcal{M}$ in the given class and every world w in the model, $\mathcal{M}, w \vDash \varphi$. More generally, $\Sigma \vDash_{\mathsf{S}} \varphi$ where Σ is a set of formulas and S is a class of models denotes that for all models $\mathcal{M}$ from S and all worlds in $\mathcal{M}$, if all formulas in Σ are satisfied on w (written as $\mathcal{M}, w \vDash \Sigma$) then $\mathcal{M}, w \vDash \varphi$.

In the case of doxastic logic, $B\varphi \to \neg B\neg\varphi$, $B\varphi \to BB\varphi$ and $\neg B\varphi \to B\neg B\varphi$, these three formulas about belief, together with axiom system K, form axiom system KD45 (Table 1.1).

Table 1.1: Axiom system KD45

Propositional Tautologies and Modus Ponens			
RE:	from $\vdash_{KD45} \varphi$ infer $\vdash_{KD45} B\varphi$		
K:	$\vdash_{KD45} B(\varphi \to \psi) \to (B\varphi \to B\psi)$	D:	$\vdash_{KD45} B\varphi \to \neg B\neg\varphi$
4:	$\vdash_{KD45} B\varphi \to BB\varphi$	5:	$\vdash_{KD45} \neg B\varphi \to B\neg B\varphi$

The symbol $\vdash_{KD45}$ in the table denotes deducibility in the axiom system KD45. With different axiom systems, the subscript varies. We use $\Sigma \vdash_{\Lambda} \varphi$ to denote the deducibility of φ from Σ in the axiom system Λ, which in this dissertation means that $\vdash_{\Lambda} \varphi$ or that there are $\psi_1, \ldots, \psi_n \in \Sigma$ such that $\vdash_{\Lambda} (\psi_1 \wedge \ldots \wedge \psi_n) \to \varphi$.

The axiom system KD45 is strongly complete with respect to the class of Kripke models whose R is serial, transitive and Euclidean, denoted by STE, which means that

$$\Sigma \vDash_{\mathsf{STE}} \varphi \quad \text{iff} \quad \Sigma \vdash_{KD45} \varphi \ .$$

A weaker form of completeness, called "weak completeness", refers to a special case of strong completeness in which $\Sigma = \varnothing$.

The concept of consistency is always associated with an axiom system. A set of formulas Σ is consistent with respect to an axiom system Λ if and only if

$\Sigma \nvdash_\Lambda \bot$, where $\bot := p \wedge \neg p$. So the hierarchy of consistency that we introduced in Section 1.1 can only be rigorously defined with respect to a specific axiom system. Take the axiom system KD45 as an example. We say an agent's belief which consists of a set of propositions Π is n-consistent with respect to KD45 if and only if for all subsets $\Sigma \subseteq \Pi$ with cardinality $l \leq n$, $\Sigma \nvdash_{KD45} \bot$; and Π is ∞-consistent if and only if for all *finite* subsets $\Sigma \subseteq \Pi$, $\Sigma \nvdash_{KD45} \bot$.

What we have just introduced is doxastic logic for a single agent. This logic can be easily extended to the multi-agent setting so that some notions of group belief can be defined, for example, distributed belief, everyone's belief and common belief. To proceed we only need to index the accessibility relation with symbols representing agents. Given a set of agents Ag, the multi-agent model is a structure (W, R_a, V) where $a \in \mathsf{Ag}$. Correspondingly, we need to index the belief operator with symbols, for example, $B_a\varphi$.

For simplicity, let's assume Ag is finite. In this setting, "everybody believes φ" can be simply defined as

$$E\varphi := \bigwedge_{a \in \mathsf{Ag}} B\varphi \ .$$

Common belief can be understood to be infinitely iterating the operator for everyone's belief, which is, nevertheless, inexpressible directly in doxastic logic. However, there is an elegant way of defining common belief in doxastic logic which is equivalent to infinitely iterating the operator E. For more details, we refer readers to (Fagin et al., 1995, Section 2.2).[6] The only point we want to bring up here is that common belief is a very strong notion in the sense that it is very hard to meet its requirement, since everyone's belief, implying a form of consensus, is already a very strong notion.

Besides everyone's belief and common belief, distributed belief is another notion of group belief studied in doxastic logic. Let R_D be a binary relation on W defined as

$$R_D = \bigcap_{a \in \mathsf{Ag}} R_a \ .$$

Distributed belief D is then define as

$$\mathcal{M}, w \vDash D\varphi \quad \text{iff} \quad \text{for all } v \in W \text{ such that } R_D wv, v \in [\![\varphi]\!]$$

Distributed belief as defined in this way can be debated as it presumes that the group can function as a single agent, collecting everyone's beliefs and closing it under logical consequences so that one can extract as much information as possible. At the same time, the here defined concept of distributed belief cannot deal with any inconsistency between different agents' beliefs. If any conflict embedded in the group members' beliefs appears, distributed belief becomes inconsistent.

[6] Although Fagin et al. (1995) is mainly about knowledge, technically, what is discussed about common knowledge in its Section 2.2 also applies to the concept of common belief in doxastic logic.

The relation between distributed belief, everyone's belief and common belief reflects the tension mentioned in the introduction. Common belief implies everyone's belief, and everyone's belief implies distributed belief. Although what constitutes everyone's belief are fully consistent, a group can rarely reach consensus. Although a group's distributed belief includes more than every group member's beliefs, it may sometimes explode because of its inconsistency.

Chapter 2

Argumentation and Belief

Once when Confucius was walking through a marketplace, he saw two children who looked like they were arguing heatedly over something. Confucius got curious and went over to ask them what their contention was.

One child said, "I say the sun is nearer us when it is rising and gets farther away at midday."

The other child immediately said, "I say the sun is farther away when it is rising and nearer us at midday."

The one who spoke first then said, "The sun looks bigger when it is at the horizon and gets smaller as it reaches noon. Don't things look smaller when they are far away and bigger when they are near?"

The second child was not daunted. He said, "The sun is hotter at noon than when it rises in the morning. Isn't something hotter when it is near and cooler when it is farther away?"

Both children then pestered Confucius to answer their questions. Confucius was stumped. He told them he couldn't tell which of them was correct.

The children laughed and said, "Hey, you're supposed to be a learned man, and you can't even answer our questions!"

(*Lieh-Tzu*)[1]

[1] The translation is from Wong (2001, Part five, Section 52)

2.1 Introduction

In the first Chapter, we mentioned that the agent's doxastic attitudes could be divided into multiple layers according to its level of consistency. Information we are receiving every day is not guaranteed to be consistent at all. The man selling shields and spears (in the story mentioned in Section 1.1.1) lived not only in ancient China but also lives in every corner of our modern world. There is no doubt that the agent's doxastic attitudes should not allow any inconsistent information of this type. However, there is no reason to stop any self-consistent information from entering the first stratum of the agent's doxastic attitudes. The agent accepts every piece of self-consistent information as evidence. Different pieces of evidence may contradict each other, but none of them should be contradictory itself.

In the quotation at the beginning of this Chapter, Confucius, one of the greatest ancient Chinese philosophers, could not decide which child is right, because he did not have any other further information which can help determine the answer. Evidence from neither child breaks through the first layer and transforms into Confucius's belief. [2]

The agent's evidence is the basis for her belief. But not every piece of the agent's evidence is eligible to be the agent's belief. Otherwise, Confucius would have believed both that "the sun is farther away when it is rising and nearer us at midday" and that "the sun is nearer us when it is rising and gets farther away at midday" simultaneously. Then under what conditions is the agent's evidence eligible to be belief? van Benthem and Pacuit (2011) initiated the research on this question from a logicians' perspective, bringing the hidden concept – evidence into the spotlight. A series of works follow this line of thought (van Benthem, Fernández-Duque and Pacuit, 2012; van Benthem, Fernández-Duque and Pacuit, 2014; Baltag, Fiutek and Smets, 2016; Baltag, Bezhanishvili, Özgün and Smets, 2016*a*; Liu and Lorini, 2016; Baltag and Occhipinti, 2017), extending and improving on the idea and the formal setting in van Benthem and Pacuit (2011). Besides these works, some other papers (Égré, Marty and Renne, 2014; Grossi and van der Hoek, 2014) also shed light on the question, although they focus more on how belief is justified by evidence or arguments rather than how belief is built upon evidence.

Given the varieties of the answers to the question proposed in the literature, our work continues and builds further on this line of research. We feature the agent's reasoning as the driving force behind the formation of belief based on evidence, which has been indicated in van Benthem and Pacuit (2011) through the study of an evidence-based notion of belief and the agent's reasoning about evidence (dynamics of evidence). The novelty of our proposal in this chapter will

[2] The story itself does not necessarily have any implication about Confucius's internal doxastic state. From the perspective of behaviourism and for our purpose of illustration, nevertheless, it should cause no confusion when we make a conjecture about Confucius's belief.

lie in the focus on a specific type of reasoning – argumentational reasoning, as it is called in Dung (1995, p.323).

> Roughly, the idea of argumentational reasoning is that a statement is believable if it can be argued successfully against attacking arguments. In other words, whether or not a rational agent believes in a statement depends on whether or not the argument supporting this statement can be *successfully defended* against the counterarguments.

Why did we suddenly start talking about arguments rather than evidence? Despite its multiple meanings, in this thesis, "argument" only refers to "a set of reasons that show that something is true or false, right or wrong". In the sense of conferring justification, it is quite similar to what evidence means (Conee and Feldman, 2004). They both refer to information accepted by the agent as her reasons for supporting something. However, "argument" emphasises more the structural property of information. [3] When we refer to one piece of evidence as an "argument", we stress that it is achieved by combining and organising different pieces of evidence in a logical way. For example, the three pieces of information given by the first child, "the sun is nearer us when it is rising and gets farther away at midday", "the sun looks bigger when it is at the horizon and gets smaller as it reaches noon", and "things look smaller when they are far away and bigger when they are near" are accepted by Confucius as evidence. By combining them, Confucius gets a new piece of evidence which is taken as an argument, supporting "the sun is nearer us when it is rising and gets farther away at midday".

Another reason to use the term "argument" is that we assume that the agent's collection of evidence can be organised as an argumentation. By "argumentation", we mean a set of arguments with attack relations between them. In an argumentation, an argument attacks or is attacked by other arguments which contradict it. For example, the arguments of the two children in the story attack each other. When Confucius heard the two arguments, he internalised them. The two arguments became a part of the argumentation happening in Confucius's mind. Although information is not always presented to the agent in the form of argumentation, nothing stops the agent from structuring her collection of evidence into an argumentation. The argumentational reasoning described in the quote from Dung (1995) can thus be employed by the agents to form her belief based on the argumentation in her mind.

After reviewing the work in van Benthem and Pacuit (2011) and Baltag, Bezhanishvili, Özgün and Smets (2016*a*) in Section 2.2, we will present the framework called "topological argumentation model" in Section 2.3, where the new notion of belief is defined based on the agent's argumentational reasoning. As the name of the framework indicates, we bring together two seemingly unrelated areas – topological semantics for evidence (Baltag, Bezhanishvili, Özgün

[3] Such an emphasis can also be found in Toulmin (1958).

and Smets, 2016*a*) and formal argumentation theory (Dung, 1995; Caminada and Gabbay, 2009; Gabbay et al., 2016). This fusion results in a new notion of belief called "grounded belief" which is a natural weakening of the topological notion of belief – justified belief – proposed in Baltag, Bezhanishvili, Özgün and Smets (2016*a*). Moreover, it differs from the latter concerning the principle of closure under conjunction, which fails for grounded belief. In the same section, we show what conditions of the attack relation can save this failure. More philosophical analysis of the failure of closure under conjunction will follow in Section 2.6, which reveals a tension between grounded belief and justified belief. Before that, Section 2.4 and Section 2.5 study respectively the logic of grounded belief and the logic where both grounded belief and the notion of "having an argument for ... " can be expressed. The logic in Section 2.5 will also play an essential role in Chapter 3, where we try to relieve the tension between grounded belief and justified belief.

2.2 Evidence Model and Topology of Evidence

In this section, we introduce the work about how the agent's belief is based on evidence in van Benthem and Pacuit (2011) and Baltag, Bezhanishvili, Özgün and Smets (2016*a*), which lay the groundwork for what we are going to develop.

In Section 2.1, we have introduced roughly the notions of evidence and argument. In this section, we will make them precise.

Logicians tend to view information as a set of possible states of affairs or possible worlds. In the sense of being informational, each piece of evidence can also be treated as a set of possible states of affairs. This is exactly what van Benthem and Pacuit (2011) do: given a non-empty set of possible worlds W, each piece of evidence accepted by the agent is modeled as a subset of W.

2.2.1. Definition ((Uniform) evidence model (van Benthem and Pacuit, 2011)). A (uniform) *evidence model* is a tuple $\mathcal{M} = (W, E_0, V)$ where

- $W \neq \varnothing$ is a set of possible worlds;
- $E_0 \subseteq 2^W - \{\varnothing\}$ is a family of non-empty subsets of W, called basic evidence collection;
- $V : \mathsf{At} \to 2^W$ is a valuation function for a given set At of atomic propositions.

The family E_0 is required to include the whole set of possible worlds W (i.e. $W \in E_0$).

Note that the evidence model we present here is actually the uniform evidence model in van Benthem and Pacuit (2011). E_0 is world-dependent in an evidence model while world-invariant in a uniform evidence model. To illustrate the idea of evidence-based belief, a uniform evidence model is sufficient. So for simplicity, we will always assume that E_0 is world-invariant.

E_0 includes all the pieces of evidence the agent acquires and accepts. We call E_0 the agent's *basic evidence collection*. Since the agent would not accept a contradiction as evidence, the empty set is not allowed in E_0. On the other hand, due to the optimism of the logicians, the agent is always armed with tautologies as her evidence, i.e., $W \in E_0$. Despite the agent's capability of doing logic, she may not be able to resolve the conflict between different pieces of evidence immediately. So there can be pieces of evidence in E_0 which contradict each other. That is, it is allowed in the evidence model that there exist $Q, Q' \in E_0$ such that $Q \cap Q' = \varnothing$.

The existence of the conflict between different pieces of evidence in the agent's basic evidence collection does not mean that the agent accepts the conflict. Instead, it provides the impetus for the agent's action of resolving the conflict. The agent has achieved some order by rejecting contradiction in the chaos of incoming information. Now the task becomes harder: she needs to resolve the conflict among pieces of evidence to achieve more harmony.

What van Benthem and Pacuit (2011) suggests is that the agent should assemble from her basic evidence collection those pieces of evidence which "fit together well", and then combine them. To be more precise, "fit together well" means it satisfies the finite intersection property.

2.2.2. Definition. Given a set of possible worlds W, a family $F \subseteq 2^W$ has the *finite intersection property* when the intersection of every finite subset of F is non-empty.

By collecting those pieces of evidence which fit together well, the agent obtains maximal bodies of evidence.

2.2.3. Definition ((Maximal) body of evidence). Let $\mathcal{M} = (W, E_0, V)$ be an evidence model. A *body of evidence* $F \subseteq E_0$ is a subfamily of E_0 which has the finite intersection property. A body of evidence is *maximal* if it cannot be properly extended to any other body of evidence.

By "combining" a maximal body of evidence F, we refer to the idea of taking the intersection $\bigcap F$. Hence through a series of collections and combinations, the agent ends up with a set

$$\{\bigcap F \subseteq W \mid F \text{ a maximal body of evidence }\} \ .$$

Note how all members in this set are in conflict with each other: given any two different maximal bodies of evidence F_1 and F_2, $\bigcap F_1 \cap \bigcap F_2 = \varnothing$. The agent manipulates her basic evidence collection logically and puts it in a logical shape. It is now clear where all the hidden conflicts among her pieces of evidence lie. The remaining task is a matter of choice – which maximal bodies of evidence should be picked.

The agent modeled by van Benthem and Pacuit (2011) is very conservative, so she decides to take all the maximal bodies of evidence into account before having any reason to reject any one of them.

2.2.4. Definition (Belief à la van Benthem and Pacuit – evidence-based belief). The agent has *evidence-based belief* of a proposition $P \subseteq W$ if and only if every maximal body of evidence $F \subseteq E_0$ supports P, i.e., $\bigcap F \subseteq P$.

What is a little surprising is that according to Definition 2.2.4 the agent may believe a contradiction, as illustrated in Example 2.2.5.

2.2.5. Example. Consider the evidence model $(\mathbb{N}, E_0 = \{[n, +\infty) \mid n \in \mathbb{N}\}, \varnothing)$. Note how E_0 itself is a body of evidence and, moreover, is the unique maximal one. But $\bigcap E_0 = \varnothing$, and thus the agent believes $\varnothing$.

The contradiction rejected at the gate of the agent's basic evidence collection sneaks into the agent's realm of beliefs. The reason for this phenomenon lies in the tension between the agent's limited ability to combine *finite* pieces of evidence when collecting maximal bodies of evidence and what is required by the notion of belief in Definition 2.2.4 – the ability to combine *infinite* pieces of evidence.

In van Benthem and Pacuit (2011), the authors also focus on logics of dynamics of evidence change, including evidence addition, evidence removal, evidence upgrade and evidence combination. Among these types of evidence dynamics, evidence combination (#) is of particular importance for understanding the later development of an evidence-based notion of belief in Baltag, Bezhanishvili, Özgün and Smets (2016*a*) and this chapter. According to its definition (van Benthem and Pacuit, 2011, Definition 4.19), by applying the operation of evidence combination, the basic evidence collection E_0 becomes $E^{\#}$ which is the smallest set closed under intersection and containing E_0. Note that $E^{\#}$ is closed under non-empty intersection rather than *finite* non-empty intersection. It reflects what is endorsed by the definition of evidence-based belief (Definition 2.2.4) about the agent's ability to combine *infinite* pieces of evidence. However, what is not shared by the definition of evidence-based belief is the requirement of the non-emptiness of the infinite intersection, which then reflects the discrepancy between the mechanism of ensuring consistency (maximal body of evidence) in evidence-based belief and the agent's infinite ability to combine evidence. There is no way of ensuring the consistency of combining infinite pieces of evidence, although the agent is endowed with the ability of infinite combination.

It is an involved issue whether the discrepancy is inherent to the internal state of any human being or not. Without touching on this knotty problem, however, just out of the temptation of logicians, consistency is taken as a desirable property. So an attempt is made to adapt the notion of evidence-based belief in van Benthem and Pacuit (2011) by Baltag, Bezhanishvili, Özgün and Smets (2016*a*) using a *topological* evidence model. Its strategy is to limit the agent's ability to combine evidence, which is embodied in the requirement of a topology that it is closed under *finite* intersection and a definition of belief based on the topology.

2.2.6. Definition (Baltag, Bezhanishvili, Özgün and Smets (2016*a*)).
A *topological evidence model* $\mathcal{M} = (W, E_0, \tau_{E_0}, V)$ extends an evidence model (W, E_0, V) (Definition 2.2.1) with τ_{E_0}, the topology over W generated by E_0. [4] For simplicity, and when no confusion arises, τ_{E_0} will be denoted simply by τ.

The elements of a topology are called *open sets*. The open sets in τ_{E_0} are all unions of *finite* intersections of elements of E_0. These unions and *finite* intersections can be seen as the agent's logical manipulation of her basic evidence. Hence each open set in τ_{E_0} is implicitly accompanied by its logical structure concerning how they are generated. Following Baltag, Bezhanishvili, Özgün and Smets (2016*a*), we call the *non-empty* open sets in τ_{E_0} "arguments". More justification for using this term can also be found in Özgün (2017, Section 5.2.2).

Take a body of evidence (Definition 2.2.3) $F \subseteq E_0$. If F is infinite, $\bigcap F$ is not necessarily an argument. If F is finite, $\bigcap F$ must be an argument. Now take a maximal body of evidence $F_{max} \subseteq E_0$. Even if $\bigcap F_{max}$ is not an argument, for every finite subset $F_{fin} \subseteq F_{max}$, $\bigcap F_{fin}$ is an argument. The notion of belief in Definition 2.2.4 leaves a fissure in the stratum of argument from which it enters into an area where the agent has no control. Therefore, to keep the notion of belief based on a solid stratum of arguments, instead of asking whether every maximal body of evidence F_{max} supports $P \subseteq W$, Baltag, Bezhanishvili, Özgün and Smets (2016*a*) asks whether each finite body of evidence can be strengthened to a certain finite body of evidence which supports P. Recall that "a body of evidence F supports P" means $\bigcap F \subseteq P$. And "strengthened to a certain finite body of evidence" means extending the body of evidence by adding more pieces of evidence while keeping it consistent.

2.2.7. Definition (Baltag, Bezhanishvili, Özgün and Smets (2016*a*)). Let $\mathcal{M}$ be a topological evidence model (W, E_0, τ, V). The agent has *justified belief* of a proposition $P \subseteq W$, denoted by $BelP$ if and only if every finite body of evidence can be strengthened to a certain finite body of evidence which supports P (i.e., for every finite body of evidence $F \subseteq E_0$ there is a finite body of evidence $F' \subseteq E_0$ such that $F \subseteq F'$ and $\bigcap F' \subseteq P$).

It turns out that this topological notion of belief satisfies the axioms of the standard doxastic logic KD45. The agent's beliefs thus achieve full consistency, which may comfort logicians a lot. Besides logicians, the topological notion of belief may also intrigue a topologist, because the topological notion of belief is, in fact, a

[4] A *topology* over a set $X \neq \varnothing$ is a family $\tau \subseteq 2^X$ containing both X and $\varnothing$, and closed under *finite* intersections and arbitrary unions. The *topology generated by a given set* $E \subseteq X$ is the smallest topology τ_E over X such that $E \subseteq \tau_E$.

purely topological notion. As Proposition 2 in Baltag, Bezhanishvili, Özgün and Smets (2016*a*) shows, $Bel P$ holds in a topological evidence model if and only if there is a dense open t in the topology of the model such that $t \subseteq P$.

Not everyone is happy, however, with the topological notion of belief. Recall the definition of dense opens in a topology.

2.2.8. Definition. Given a topological space (X, τ) where τ is a topology over the set X, $t \in \tau$ is a *dense open* if and only if it intersects with all the other nonempty opens in τ.

Thus the equivalence shown in Proposition 2 of Baltag, Bezhanishvili, Özgün and Smets (2016*a*) implies that the topological notion of belief requires what the agent believes should be supported by an argument which is consistent with any other arguments. [5] Such a requirement indicates that every argument is equally important and no argument should be ignored when deciding what to believe. The possible consequence is that the agent is too conservative to believe anything. Our work in the next section tries to make the agent bolder.

2.3 Belief Grounded on Arguments

2.3.1 Topological argumentation model

Let us get back to the example of Confucius (on page 15). Confucius got two arguments which attack each other. Now consider a modern astronomer who is reading this story about the debate between the two children. Of course, she knows the right answer to the question. Moreover, she has a scientific argument which supports her answer. Denote the argument for the right answer in the astronomer's mind by A, the first child's argument by B; and the second child's argument by C. In the mind of the astronomer, B and C attack each other and A attacks both B and C, but neither B nor C attacks A.

For an agent, it is possible that some arguments in her mind defeat some others, just as in the case of the astronomer. There is no reason for always treating all the arguments on a par. The topological notion of belief, however, never discriminates in favour of any side when two arguments conflict with each other,[6] in part because the topological evidence model does not provide further information about how the agent evaluates her arguments.

To equip the model with the ability of specifying the attack relation between different arguments as in the mind of the astronomer, we add an "attack" relation $\leftarrowtail$ to the topological evidence model:

[5] We will also use "an argument $t \in \tau$ supports P" to express $t \subseteq P$.

[6] In a topological evidence model, two arguments $t, t' \in \tau$ conflict with each other if and only if $t \cap t' = \varnothing$.

2.3.1. Definition (Topological argumentation model). A *topological argumentation model* is a tuple $\mathcal{M} = (W, E_0, \tau_{E_0}, \twoheadleftarrow, V)$ which extends a topological evidence model (Definition 2.2.6) with a relation $\twoheadleftarrow \subseteq (\tau \times \tau)$, called the *attack relation* (with $t_1 \twoheadleftarrow t_2$ read as "t_2 *attacks* t_1"), which is required to satisfy the following three conditions:

1. for all $t_1, t_2 \in \tau$: $t_1 \cap t_2 = \varnothing$ if and only if $t_1 \twoheadleftarrow t_2$ or $t_2 \twoheadleftarrow t_1$;
2. for all $t, t_1, t_1', \in \tau$: if $t_1 \twoheadleftarrow t$ and $t_1' \subseteq t_1$, then $t_1' \twoheadleftarrow t$;
3. for all $t \in \tau \setminus \{\varnothing\}$: $\varnothing \twoheadleftarrow t$ and $t \not\twoheadleftarrow \varnothing$.

The three conditions imposed on the attack relation $\twoheadleftarrow$ are self-explanatory. Due to the first condition, the model can deal with the case of the astronomer, for whom the attack relation is not symmetric. The second condition says that if t attacks t_1, then it should also attack any stronger arguments. The last condition confirms the priority of all arguments over a contradiction.

In a topological argumentation model, the topology τ, together with the attack relation $\twoheadleftarrow$, represents the argumentation happening in the agent's mind. We denote it by $\mathcal{A}_\tau = (\tau, \twoheadleftarrow)$ and call it "the agent's argumentation framework". Next we address the question: how the agent reasons in her argumentation framework to form her belief.

2.3.2 Grounded belief

To answer the above question, recall the quote on page 17 and pay attention to a keyword in it – successfully defended. According to Dung, a statement is believable if and only if there is an argument supporting the statement which can be successfully defended against any other counterarguments. In Dung (1995), he proposes several different ways of formalising the idea of "successfully defended", as grounded semantics, preferred semantics and stable semantics. Here, we adopt grounded semantics.

Let us first illustrate the intuitive idea behind the formal definition of the defending relation in an agent's argumentation framework before we present it formally.

2.3.2. Example. Consider the topological argumentation model

$$\mathcal{M}_{123} = (\{1,2,3\}, E_0 = \{\{1\},\{2\},\{3\}\}, \tau = 2^{\{1,2,3\}}, \twoheadleftarrow, V) \tag{2.1}$$

where $\{1\} \twoheadleftarrow \{2\}, \{1\} \twoheadleftarrow \{3\}, \{1\} \twoheadleftarrow \{2,3\}, \{2\} \twoheadleftarrow \{1,3\}, \{3\} \twoheadleftarrow \{2\}, \{3\} \twoheadleftarrow \{1\}$, $\{3\} \twoheadleftarrow \{1,2\}$ and $\{1,3\} \twoheadleftarrow \{2\}$, and the empty open is attacked by all opens in τ, as shown in Figure 2.1 where $\twoheadleftarrow$ is draw as a plain arrow without the tail. Alternatively, we can draw the agent's argumentation framework for $\mathcal{M}_{123}$ as a graph (Figure 2.2). Note that the attack relations involving the empty open are not drawn.

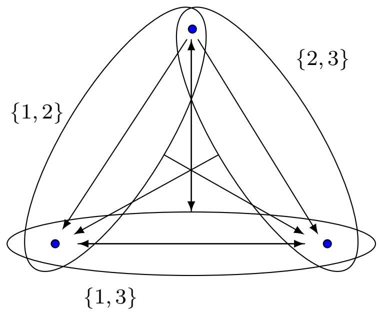

Figure 2.1: The topological argumentation model $\mathcal{M}_{123}$

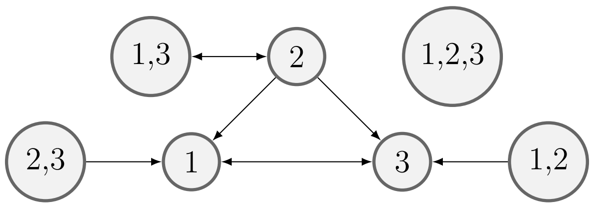

Figure 2.2: The agent's argumentation framework for $\mathcal{M}_{123}$

From the graph in Figure 2.2, we can see that argument $\{1\}$ is attacked by argument $\{3\}$. Meanwhile, $\{3\}$ is attacked by $\{1,2\}$. In this case, we say that argument $\{1,2\}$ defends argument $\{1\}$ from its counterargument $\{3\}$. For the same reason, $\{1,3\}$ defends $\{1\}$ from its another counterargument $\{2\}$. Do arguments $\{1,2\}$ and $\{1,3\}$ together defend $\{1\}$? No, because they do not defend $\{1\}$ from the attack of $\{2,3\}$. The intuition is that for an argument to be defended by a set of arguments, the set of arguments has to defend the argument from all of its counterarguments. This intuition leads us to the following formal definition of defending in a topological argumentation model.

2.3.3. Definition (Defending). Given a topological argumentation model $\mathcal{M}$, an argument $t \in \tau$ is *defended by a set of arguments* $X \subseteq \tau$ if and only if for all arguments $y \in \tau$ such that $t \twoheadleftarrow y$, there is an argument $x \in X$ such that $y \twoheadleftarrow x$.

According to this definition, $\{1\}$ in Figure 2.2 is not defended by any set of arguments , and $\{2\}$ is defended by itself. Then can $\{2\}$ be defended successfully according to grounded semantics?

The following function is key to the introduction of grounded semantics.

2.3.4. Definition (Characteristic function). Given a topological argumentation model $\mathcal{M}$, the *characteristic function* $d_{\mathcal{M}} : 2^{\tau} \mapsto 2^{\tau}$ is defined as follows:

$$d_{\mathcal{M}}(X) = \{t \in \tau \mid t \text{ is defended by } X\}$$

where $X \subseteq \tau$.

Note that the characteristic function d is monotonic. Thus, since the power set of τ is a complete lattice, the least fixed point of d can be built up from the empty set, according to the classical results about the fixed points of a monotonic function on a complete lattice (Arnold and Niwiński, 2001, Theorem 1.2.8 (Knaster–Tarski theorem) and Theorem 1.2.11).

The building procedure can be described as follows. The first step is $X_1 = d(\varnothing)$, including those arguments which can be defended by an empty set. In other words, these arguments are not attacked at all. Note that dense opens are all in X_1. In the second step, the collected arguments are $X_2 = d(X_1)$. In the $(\alpha+1)$th step, the collected arguments are $X_{\alpha+1} = d(X_\alpha)$. For every limit ordinal β, $X_\beta = \bigcup_{\alpha\leq\beta} X_\alpha$. Note that for all ordinals α and α' such that $\alpha \leq \alpha'$, $X_\alpha \subseteq X_{\alpha'}$, because d is a monotonic function. Thus there is an ordinal α in the procedure such that X_α is the least fixed point of the function d. [7] This implies that for all β such that $\beta \geq \alpha$, $X_\beta = X_\alpha$, which means that the procedure of accumulation stops increasing at some point, say X_α, and this is the least fixed point of the function d.

The existence of such a stopping point of the procedure, which turns out to be the least fixed point of the characteristic function d, ensures a well-defined notion of belief as follows:

2.3.5. Definition (Grounded belief). Let $\mathcal{M}$ be the topological argumentation model $(W, E_0, \tau_{E_0}, \twoheadleftarrow, V)$. The agent has *grounded belief* of a proposition $P \subseteq W$ (notation: $\mathcal{B}P$) if and only if there is an open set in LFP_τ supporting P, that is

$$\mathcal{B}P \quad \text{if and only if} \quad \exists\, \mathsf{f} \in \mathsf{LFP}_\tau : \mathsf{f} \subseteq P$$

where LFP_τ denotes the least fixed point of the characteristic function $d_{\mathcal{M}}$.

2.3.3 Logical properties of grounded belief and its relation to justified belief

The readers who are familiar with formal argumentation theory (Dung, 1995) may realise that the above way in which we have defined grounded belief is nothing but an application of the grounded semantics in Dung (1995) to the topological space. Hence it inherits all the nice properties of the grounded semantics. For example,

2.3.6. Proposition (Conflict-free of LFP_τ). *Given a topological argumentation model $\mathcal{M}$, for all* $\mathsf{f}, \mathsf{f}' \in \mathsf{LFP}_\tau$,

$$\mathsf{f} \cap \mathsf{f}' \neq \varnothing \ .$$

[7] The least fixed point of d is a set of arguments X such that $X = d(X)$ and $X = \bigcap_{Y=d(Y)} Y$, i.e., the least set satisfying $Y = d(Y)$

Proof:
Let $X_0 = \varnothing$, $X_{\alpha+1} = d(X_\alpha)$ and when β is a limit ordinal $X_\beta = \bigcup_{\alpha<\beta} X_\alpha$. The proof proceeds by a complete induction on the ordinal β, since $\varnothing$ is conflict-free and if X_β is conflict-free then $X_{\beta+1}$ is conflict-free □

The conflict-free property of LFP_τ ensures the property of mutual consistency for the agent's grounded beliefs.

2.3.7. COROLLARY. *Given a topological argumentation model $\mathcal{M}$, if $\mathcal{B}P$ holds, then for all $Q \subseteq W$ such that $Q \cap P = \varnothing$, $\mathcal{B}Q$ does not hold.*

Besides the mutual consistency, the agent's grounded beliefs are upward-closed ($\mathcal{B}P$ and $P \subseteq Q$ imply $\mathcal{B}Q$), which is implied directly by the definition of grounded belief. In fact, a stronger claim also holds: LFP_τ itself is closed upwards.

2.3.8. PROPOSITION. *Given a topological argumentation model, if $\mathsf{f} \in \mathsf{LFP}_\tau$ and $\mathsf{f}' \in \tau$ satisfies that $\mathsf{f} \subseteq \mathsf{f}'$, then $\mathsf{f}' \in \mathsf{LFP}_\tau$.*

Proof:
Take any $\mathsf{f} \in \mathsf{LFP}_\tau$ and any $\mathsf{f}' \in \tau$ such that $\mathsf{f} \subseteq \mathsf{f}'$. Suppose no one attacks f'; then we are done as, by LFP_τ's definition, every non-attacked element of τ should be in LFP_τ. Suppose otherwise, and let t be one of such opens attacking f'; it is enough to find an $\mathsf{f}'' \in \mathsf{LFP}_\tau$ attacking t, as then f' would be defended by f'' and thus, by definition, f' would be in LFP_τ. Now, since t attacks f', it should also attack the stronger f (as required by $\twoheadleftarrow$'s definition); but then, since f is in LFP_τ, it should be defended by someone in LFP_τ, that is, there is a f'' in LFP_τ attacking t. This completes the proof. □

Are the agent's grounded beliefs closed under conjunction? No. Example 2.3.2 is a counterexample. $\mathsf{LFP}_\tau = \{\{1,2\},\{2,3\},\{1,2,3\}\}$ for the topological argumentation model in 2.1, but $\{2\} \notin \mathsf{LFP}_\tau$.

We will discuss the failure of closure under conjunction in Section 2.6 in more detail. For now, let us focus on the possible reasons behind the failure of this closure property, by identifying additional conditions under which the property holds.

2.3.9. PROPOSITION. *Let $\mathcal{M}$ be a topological argumentation model and LFP_τ the least fixed point for the characteristic function $d_\mathcal{M}$.*

- *If $\twoheadleftarrow$ is unambiguous (i.e. for all $t_1, t_2, t_3 \in \tau$, if $t_1 \twoheadleftarrow t_2$ and $t_2 \twoheadleftarrow t_3$, then $t_1 \not\twoheadleftarrow t_3$ and $t_3 \not\twoheadleftarrow t_1$), then LFP_τ is closed under intersections;*
- *If $\twoheadleftarrow$ is symmetric on the set of arguments(i.e. for all $t_1, t_2 \in \tau \setminus \{\varnothing\}$, if $t_1 \twoheadleftarrow t_2$ then $t_2 \twoheadleftarrow t_1$), then LFP_τ is closed under intersections.*

Proof:
The following lemma will be useful in this proof.

2.3.10. Lemma. *Let $\mathcal{M} = (W, E_0, \tau_{E_0}, \twoheadleftarrow, V)$ be a topological argumentation model. Then, for all $\mathsf{f}_1, \mathsf{f}_2 \in \mathsf{LFP}_\tau$, $\mathsf{f}_1, \mathsf{f}_2 \in \mathsf{LFP}_\tau$ implies $\mathsf{f}_1 \cap \mathsf{f}_2 \in \mathsf{LFP}_\tau$ if and only if for all $t \in \tau$, if $\mathsf{f}_1 \cap \mathsf{f}_2 \twoheadleftarrow t$, then $t \cap \mathsf{f} = \varnothing$ for some $\mathsf{f} \in \mathsf{LFP}_\tau$.*

Proof:
From left to right, take arbitrary $\mathsf{f}_1, \mathsf{f}_2 \in \mathsf{LFP}_\tau$. Suppose there is an open $t \in \tau$ such that t attacks $\mathsf{f}_1 \cap \mathsf{f}_2$ but is not in conflict with anybody in LFP_τ. From the latter it follows that nobody in LFP_τ attacks t, and thus the attacked $\mathsf{f}_1 \cap \mathsf{f}_2$ is not defended by LFP_τ; therefore, $\mathsf{f}_1 \cap \mathsf{f}_2$ is not in LFP_τ.

From right to left, take arbitrary $\mathsf{f}_1, \mathsf{f}_2 \in \mathsf{LFP}_\tau$ and take an argument $t \in \tau$ such that $\mathsf{f}_1 \cap \mathsf{f}_2 \twoheadleftarrow t$. It implies that there is $\mathsf{f}' \in \mathsf{LFP}_\tau$ such that $t \cap \mathsf{f}' = \varnothing$. Thus either f' attacks t or else t attacks f'. The former case implies that there is $\mathsf{f}'' \in \mathsf{LFP}_\tau$ such that $t \twoheadleftarrow \mathsf{f}''$ by virtue of f''s membership in LFP_τ; together with the latter case, i.e. $t \twoheadleftarrow \mathsf{f}'$, we can conclude that there is $\mathsf{f} \in \mathsf{LFP}_\tau$ such that $t \twoheadleftarrow \mathsf{f}$. Hence for all $t \in \tau$ such that $\mathsf{f}_1 \cap \mathsf{f}_2 \twoheadleftarrow t$, there is $\mathsf{f} \in \mathsf{LFP}_\tau$ such that $t \twoheadleftarrow \mathsf{f}$, which implies that $\mathsf{f}_1 \cap \mathsf{f}_2 \in \mathsf{LFP}_\tau$. □

Now, for Proposition 2.3.9. The proof for transitivity proceeds by contraposition. So take any $\mathsf{f}_1, \mathsf{f}_2 \in \mathsf{LFP}_\tau$ such that $\mathsf{f}_1 \cap \mathsf{f}_2$ is not in LFP_τ. Then, by Lemma 2.3.10, there is an open $t \in \tau$ who attacks $\mathsf{f}_1 \cap \mathsf{f}_2$ (i.e., $\mathsf{f}_1 \cap \mathsf{f}_2 \twoheadleftarrow t$) and who is not in conflict with elements of LFP_τ (i.e., $\mathsf{f} \in \mathsf{LFP}_\tau$ implies $t \cap \mathsf{f} \neq \varnothing$). The goal is to show that $\twoheadleftarrow$ is not unambiguous.

Define t_1, t_2 and t_3 as

$$t_1 := \mathsf{f}_1 \cap t, \qquad t_2 := \mathsf{f}_2 \cap t, \qquad t_3 := \mathsf{f}_1 \cap \mathsf{f}_2,$$

and note that none of them are empty. Note also that due to the fact that t attacks $\mathsf{f}_1 \cap \mathsf{f}_2$ ($\mathsf{f}_1 \cap \mathsf{f}_2 \twoheadleftarrow t$), t must be in conflict with $\mathsf{f}_1 \cap \mathsf{f}_2$ ($(\mathsf{f}_1 \cap \mathsf{f}_2) \cap t = \varnothing$); hence, $t_1 \cap t_2 = t_2 \cap t_3 = t_3 \cap t_1 = \varnothing$.

For **unambiguity**, consider two cases, $t_1 \twoheadleftarrow t_2$ or $t_2 \twoheadleftarrow t_1$. In the case of $t_1 \twoheadleftarrow t_2$, if $t_2 \twoheadleftarrow t_3$, then no matter $t_1 \twoheadleftarrow t_3$ or $t_3 \twoheadleftarrow t_1$, it is not unambiguous. If $t_3 \twoheadleftarrow t_2$, similarly no matter if $t_1 \twoheadleftarrow t_3$ or $t_3 \twoheadleftarrow t_1$, $\twoheadleftarrow$ is not unambiguous. The proof for the case of $t_2 \twoheadleftarrow t_1$ follows a similar argument. Therefore, we conclude that $\twoheadleftarrow$ cannot be unambiguous.

For **symmetry**, assume that $\twoheadleftarrow$ is symmetric on the set of arguments. Observe that $\mathsf{LFP}_\tau = \{t \in \tau \mid \forall x \in \tau \setminus \{\varnothing\} : x \cap t \neq \varnothing\}$ which is closed under conjunction. □

The condition "unambiguity" tells us that the failure of closure under conjunction may be caused by the existence of some argument being both a defender and an attacker of another argument. The condition "symmetry on the set of all arguments" is interesting not only because of saving the failure of closure but also in the sense of building up the connection between grounded belief (Definition 2.3.5) and justified belief (Definition 2.2.7).

In the proof of Proposition 2.3.9 for the case of symmetry, we show that when $\twoheadleftarrow$ is symmetric on $\tau \setminus \{\varnothing\}$, $\mathsf{LFP}_\tau = \{t \in \tau \mid \forall x \in \tau : x \cap t \neq \varnothing\}$. Recall that the justified beliefs can be equivalently defined as those propositions supported by some dense open in τ. Observe that

$$\{t \in \tau \mid \forall x \in \tau \setminus \{\varnothing\} : x \cap t \neq \varnothing\} = \{t \in \tau \mid t \text{ is a dense open}\} \ .$$

So when $\twoheadleftarrow$ is symmetric on the set of all arguments, grounded belief and justified belief are equivalent.

2.3.11. PROPOSITION. *Given a topological argumentation model $\mathcal{M}$, if the attack relation $\twoheadleftarrow$ is symmetric on the set of arguments $\tau \setminus \{\varnothing\}$, then $Bel\,P = \mathcal{B}\,P$.*

Moreover, the following proposition and corollary reveal the general relation between grounded belief and justified belief.

2.3.12. PROPOSITION. *Given any topological argumentation model $\mathcal{M}$,*

$$\{t \in \tau \mid t \textit{ is a dense open}\} \subseteq d_\tau(\varnothing) \ .$$

2.3.13. COROLLARY. *Given any topological argumentation model $\mathcal{M}$, if $Bel\,P$ holds, then $\mathcal{B}\,P$ holds.*

Therefore, grounded belief allows more propositions to be believed by the agent than justified belief. The agent becomes less conservative concerning what she would believe.

Does grounded belief make the agent too bold? After all, it tolerates inconsistency, although it forbids mutual inconsistency.

2.3.14. EXAMPLE. Take Example 2.3.2 and modify the attack relation a little by removing attack relation from $\{2\}$ to $\{1,3\}$. We get a new topological argumentation model whose LFP_τ becomes

$$\{\{1,2\},\{2,3\},\{3,1\},\{1,2,3\}\} \ .$$

Obviously, this set is not consistent in the sense of resulting in an empty set by taking the intersection of all its members.

We will discuss this point in Section 2.6. Before that, we need a further study of the logic of grounded belief.

2.4 The Logic of Grounded Belief

In this section, we study the logic of the agent's grounded belief (GBL). As we will show, several simple axioms soundly and completely characterise the logic of grounded belief.

The standard language of doxastic logic serves as the syntax. To emphasise that the belief operator now denotes the agent's grounded belief, we use the typeface $\mathcal{B}$ as in the previous section.

2.4.1. DEFINITION (Language of GBL $\mathcal{L}$). Let At be the set of atomic propositions. The language of GBL $\mathcal{L}$ is generated by the following grammar.

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathcal{B}\varphi$$

where $p \in \mathsf{At}$.

$\mathcal{B}\varphi$ reads as "the agent has a grounded belief of φ". All the discussions of the semantic definition of the notion of grounded belief in the topological argumentation model are crystallised in Definition 2.3.5, which leads us to the following semantics of its logic.

2.4.2. DEFINITION (Semantics of GBL). Given a topological argumentation model $\mathcal{M} = (W, E_0, \tau_{E_0}, \twoheadleftarrow, V)$ and a possible world $w \in W$, the truth conditions of formulas in $\mathcal{L}$ are defined as follows,

$\mathcal{M}, w \vDash p$	iff	$w \in V(p)$
$\mathcal{M}, w \vDash \neg\varphi$	iff	$\mathcal{M}, w \nvDash \varphi$
$\mathcal{M}, w \vDash \varphi \wedge \psi$	iff	$\mathcal{M}, w \vDash \varphi$ and $\mathcal{M}, w \vDash \psi$
$\mathcal{M}, w \vDash \mathcal{B}\varphi$	iff	there exists $\mathsf{f} \in \mathsf{LFP}_\tau$ such that $\mathsf{f} \subseteq [\![\varphi]\!]_{\mathcal{M}}$

where $[\![\varphi]\!]_{\mathcal{M}} := \{x \in W \mid \mathcal{M}, x \vDash \varphi\}$ is the set of possible worlds satisfying φ in $\mathcal{M}$ and LFP_τ is the least fixed point of the characteristic function $d_{\mathcal{M}}$.

The validity of a formulas in $\mathcal{L}$ with respect to a class of topological argumentation models is defined in a standard way as introduced in Section 1.2.

We have seen two valid formulas in GBL when discussing the logical properties of grounded belief in Section 2.3.3.

2.4.3. FACT.

$$\vDash \mathcal{B}\varphi \to \neg\mathcal{B}\neg\varphi \qquad \vDash \mathcal{B}(\varphi \wedge \psi) \to \mathcal{B}\varphi \wedge \mathcal{B}\psi$$

In addition, by noticing that the truth condition of $\mathcal{B}\varphi$ is not world-dependent, we can easily get the following two valid formulas about grounded belief.

2.4.4. FACT.

$$\vDash \mathcal{B}\varphi \to \mathcal{B}\mathcal{B}\varphi \qquad \vDash \neg\mathcal{B}\varphi \to \mathcal{B}\neg\mathcal{B}\varphi$$

They indicate that the agent is fully introspective with respect to her grounded beliefs. Furthermore, it is worth bringing up an implicit assumption about the agent – logical omniscience. It is embodied in the following two valid statements.

2.4.5. FACT.

$$\vDash \mathcal{B}\top \qquad \text{if } \vDash \varphi \leftrightarrow \psi \text{ then } \vDash \mathcal{B}\varphi \leftrightarrow \mathcal{B}\psi$$

The fist validity is due to the fact that $W \in \mathsf{LFP}_\tau$. And the second validity is due to the way in which $\mathcal{B}\varphi$ is defined.

Are these valid statement complete for axiomatising the logic? Yes. They constitute the complete axiom system for the logic of grounded belief. The answer may seem surprising because the axiom system is relatively simple compared to how grounded belief is defined semantically. It may be a consequence of the oversimplification of the language, which is not expressive enough to convey all the delicacies in the interaction between belief and arguments. Furthermore, considering that the justified belief in Baltag, Bezhanishvili, Özgün and Smets (2016*a*), as a special case of the grounded belief (Proposition 2.3.11 and Corollary 2.3.13), is completely axiomatized by the system KD45, it is not so surprising that the grounded belief is completely axiomatized by the system EMND45 (see Table 2.1). According to Chellas (1980, Figure 8.1 on p.237), axiom system EMN is obtained by taking away the axiom schema C from system EMCN, which is equivalent to system K. So EMND45 can be understood as the system KD45 without the axiom schema C: $(\mathcal{B}\varphi \wedge \mathcal{B}\psi) \to \mathcal{B}(\varphi \wedge \psi)$. Note that the monotonicity rule denoted by "RM" in Chellas (1980)

$$\text{from } \varphi \to \psi \text{ iner } \mathcal{B}\varphi \to \mathcal{B}\psi$$

is derivable from axiom M and vice versa, when RE is in the axiom system.

Table 2.1: Axiom system EMND45 for GBL

Propositional Tautologies and Modus Ponens	
4: $\mathcal{B}\varphi \to \mathcal{B}\mathcal{B}\varphi$	5: $\neg\mathcal{B}\varphi \to \mathcal{B}\neg\mathcal{B}\varphi$
RE: from $\varphi \leftrightarrow \psi$ infer $\mathcal{B}\varphi \leftrightarrow \mathcal{B}\psi$	D: $\mathcal{B}\varphi \to \neg\mathcal{B}\neg\varphi$
M: $\mathcal{B}(\varphi \wedge \psi) \to (\mathcal{B}\varphi \wedge \mathcal{B}\psi)$	N: $\mathcal{B}\top$

It is well-known that the axiom system EMN is sound and complete with respect to the class of neighbourhood models that are *supplemented* (i.e., closed under supersets) and contain the unit (i.e., the domain is in the neighbourhood) (Chellas, 1980). Since the axiom system EMND45 extends the axiom system EMN with axioms 4, 5 and D, it suggests a detour for proving the completeness result for the logic with respect to topological argumentation models.

2.4.6. THEOREM. *For all $\varphi \in \mathcal{L}$ and all $\Phi \subseteq \mathcal{L}$,*

$$\Phi \vdash_{\text{EMND45}} \varphi \qquad \textit{if and only if} \qquad \Phi \vDash \varphi \ .$$

Proof:
See Section A.1. □

Although we can see clearly how the agent's grounded beliefs behave logically, the syntax of the GBL does not exhaust the richness of the topological argumentation model. In the next section, we turn to a more expressive language in which we can also talk about the agent's arguments.

2.5 The Logic of Argument and Belief

This section presents the logic of argument and belief (abbreviated as ABL) where we can express "the agent has an argument for ..." and "the agent has grounded belief of ...". It provides us with the capability of reasoning about the relationship between these two notions in this logic. The challenge of axiomatization also comes with the more expressive syntax. They are the two sides of the same coin.

2.5.1. DEFINITION. Let $\mathsf{At} = \{p, q, r, \ldots\}$ be a set of atomic propositions. The language $\mathcal{L}_{\forall\mathcal{T}\Box}$ of ABL is generated by the following grammar:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box\varphi \mid \forall\varphi \mid \mathcal{T}\varphi$$

where $p \in \mathsf{At}$.

$\Box$ is the operator for *factive* combined evidence in Baltag, Bezhanishvili, Özgün and Smets (2016*a*), where

$$E = \{\bigcap F \subseteq W \mid F \text{ is a finite body of evidence }\}$$

is called the set of combined evidence (see Definition 2.2.3 for the definition of the body of evidence) and the pieces of *factive* combined evidence are those belonging to E and including the actual world. $\Box\varphi$ is read as "the agent has factive combined evidence for φ". $\forall$ is the universal modality. Rather than an operator for grounded belief, the last operator $\mathcal{T}$ is the operator for "belief grounded on factive combined evidence". Compared to the language $\mathcal{L}_{\Box\forall}$ studied in Baltag, Bezhanishvili, Özgün and Smets (2016*a*), we just add the operator $\mathcal{T}$. We will show later (Proposition 2.5.3) that it is equivalent to start with a language $\mathcal{L}_{\forall\mathcal{B}\Box}$ including the operator $\mathcal{B}$ directly for "the agent has grounded belief of ..." rather than $\mathcal{T}$. The reason for choosing $\mathcal{T}$ rather than $\mathcal{B}$ is that it facilitates the axiomatization and the completeness proof.

As shown in Baltag, Bezhanishvili, Özgün and Smets (2016*a*), the topological notion of belief – justified belief Bel – can be expressed in $\mathcal{L}_{\Box\forall}$:

$$Bel\,\varphi := \forall \Diamond \Box\varphi$$

where $\Diamond := \neg \Box \neg$ is the dual of $\Box$.

Similarly, "the agent has an argument for ..." and "the agent has grounded belief of ..." can be defined syntactically in $\mathcal{L}_{\forall \mathcal{T} \Box}$:

$$\boxdot\varphi := \exists \Box \varphi \qquad \mathcal{B}\varphi := \exists \mathcal{T} \varphi$$

where $\boxdot\varphi$ is the operator for "the agent has an argument for ..." and $\exists := \neg\forall\neg$ is the dual of $\forall$. To get a more intuitive understanding of these two syntactical definitions of the operator $\boxdot$ and $\mathcal{B}$, let us turn to the semantics of the logic.

2.5.2. Definition. Given a topological argumentation model $\mathcal{M}$ and a possible world w in it, the truth conditions of formulas in $\mathcal{L}_{\forall \mathcal{T} \Box}$ is defined as follows:

$\mathcal{M}, w \models p$	iff	$w \in V(p)$
$\mathcal{M}, w \models \neg\varphi$	iff	$\mathcal{M}, w \not\models \varphi$
$\mathcal{M}, w \models \varphi \wedge \psi$	iff	$\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$
$\mathcal{M}, w \models \Box\varphi$	iff	there exists $e \in E$ such that $w \in e$ and $e \subseteq [\![\varphi]\!]$
$\mathcal{M}, w \models \mathcal{T}\varphi$	iff	there exists $\mathsf{f} \in \mathsf{LFP}$ such that $w \in \mathsf{f}$ and $\mathsf{f} \subseteq [\![\varphi]\!]$
$\mathcal{M}, w \models \forall\varphi$	iff	$W \subseteq [\![\varphi]\!]$

In the truth conditions of $\Box$ and $\mathcal{T}$, the term "factive" is reflected by the membership of the actual world w.

Recall that the truth condition of $\mathcal{B}$ is "there exists $\mathsf{f} \in \mathsf{LFP}_\tau$ such that $\mathsf{f} \subseteq [\![\varphi]\!]_\mathcal{M}$" (Definition 2.4.2). Compare it with the truth condition of $\mathcal{T}$. It is straightforward to see that $\exists \mathcal{T} \varphi$ has precisely the same semantic meaning as $\mathcal{B}$.

The same observation applies to $\boxdot$. The only thing we need to note is the relationship between "having combined evidence for .. " and "having an argument for ...". The operator $\boxdot$ semantically means that there is a piece of combined evidence $e \in E$ such that $e \subseteq [\![\varphi]\!]$. This is semantically equivalent to the condition that there is an argument $t \in \tau \setminus \{\varnothing\}$ such that $t \subseteq [\![\varphi]\!]$. Hence we interpret $\boxdot\varphi$ as "the agent has an argument for φ".

The next valid formula shows that $\mathcal{T}$ is also definable by $\mathcal{B}$ and $\Box$:

2.5.3. Proposition.

$$\models \mathcal{T}\varphi \leftrightarrow (\mathcal{B}\varphi \wedge \Box\varphi) \tag{2.2}$$

Proof:
From left to right, the proof is trivial. The critical fact for proving the direction from right to left is the closure under superset of the least fixed point LFP_τ. $\Box$

According to Baltag, Bezhanishvili, Özgün and Smets (2016*a*, Theorem 4), the following axiom system is sound and complete for the language $\mathcal{L}_{\Box\forall}$ with respect to the class of topological evidence models (Definition 2.2.1):

1. propositional tautologies and Modus Ponens;

2. the S5 axioms and rules for $\forall$;
3. The S4 axioms and rules for $\Box$;
4. $\forall\varphi \to \Box\varphi$.

Now, ABL extends the language $\mathcal{L}_{\Box\forall}$ with the operator $\mathcal{T}$, and extends the topological evidence model to the topological argumentation by adding the attack relation $\leftarrowtail$. Then what is the sound and complete axiom system for ABL?

Next, we present some formulas whose validity will make them the building blocks of the axiom system. They embody crucial connections between $\Box$, $\mathcal{T}$, $\boxdot$ and $\mathcal{B}$. The whole axiom system denoted by "ABS" can be found in Table 2.2

Table 2.2: Axiom system ABS for ABL

Propositional tautologies and Modus Ponens	
The S5 axioms and rules for $\forall$	
The S4 axioms and rules for $\Box$	
$\mathcal{T}\varphi \to \mathcal{T}\mathcal{T}\varphi$	$\mathcal{T}\varphi \to \varphi$
From $\varphi \to \psi$ infer $\mathcal{T}\varphi \to \mathcal{T}\psi$	$\mathcal{T}\top$
$(\mathcal{T}\varphi \wedge \forall\psi) \to \mathcal{T}(\varphi \wedge \forall\psi)$	$\mathcal{B}\varphi \to \neg\mathcal{B}\neg\varphi$
$\mathcal{B}\varphi \wedge \neg\mathcal{B}\psi \wedge \forall((\varphi \wedge \psi) \to \Box(\varphi \wedge \psi)) \to \boxdot(\varphi \wedge \neg\psi)$	$\mathcal{T}\varphi \to \Box\varphi$
$\mathcal{T}\varphi \to \forall(\Box\varphi \to \mathcal{T}\varphi)$	$\forall\Diamond\Box\varphi \to \mathcal{B}\varphi$

2.5.4. Proposition.

$$\vDash \mathcal{B}\varphi \wedge \neg\mathcal{B}\psi \wedge \forall((\varphi \wedge \psi) \to \Box(\varphi \wedge \psi)) \to \boxdot(\varphi \wedge \neg\psi) \ .$$

Proof:

Given a topological argumentation model $\mathcal{M}$, assume that, in a possible world w, the agent has grounded belief of φ but does not have grounded belief of ψ ($\mathcal{M}, w \vDash \mathcal{B}\varphi \wedge \neg\mathcal{B}\psi$). Moreover, assume that the agent has argument $[\![\varphi \wedge \psi]\!]_{\mathcal{M}} \in \tau$[8], namely

$$\mathcal{M}, w \vDash \forall((\varphi \wedge \psi) \to \Box(\varphi \wedge \psi)) \ .$$

Putting these clues together, we have

$$\mathcal{M}, w \vDash \mathcal{B}\varphi \wedge \neg\mathcal{B}\psi \wedge \forall((\varphi \wedge \psi) \to \Box(\varphi \wedge \psi)) \ .$$

The agent has grounded belief of φ in w. It implies that the agent has an argument $\mathsf{f} \in \mathsf{LFP}_\tau$ such that $\mathsf{f} \subseteq [\![\varphi]\!]$. The agent has an argument $[\![\varphi \wedge \psi]\!]_{\mathcal{M}} \in \tau$ and does not has grounded belief of ψ. They together imply that the argument $[\![\varphi \wedge \psi]\!]$ cannot be in LFP_τ.

[8] Note the difference between "the agent has argument for $[\![\varphi \wedge \psi]\!]_{\mathcal{M}}$" and "the agent has argument $[\![\varphi \wedge \psi]\!]_{\mathcal{M}} \in \tau$. The former is expressed by $\boxdot(\varphi \wedge \psi)$.

$[\![\varphi \wedge \psi]\!] \notin \mathsf{LFP}_\tau$ implies that there must be another argument, denoted by t, attacking $[\![\varphi \wedge \psi]\!]$ but not attacked by any argument in LFP_τ. Because t is not attacked by any argument in LFP_τ, $t \cap \mathsf{f}$ is nonempty and hence an argument. $[\![\varphi \wedge \psi]\!] \twoheadleftarrow t$ tells us that $[\![\varphi \wedge \psi]\!] \cap t = \varnothing$. Therefore $\mathsf{f} \cap t \subseteq [\![\varphi]\!] \cap (W \setminus [\![\varphi \wedge \psi]\!]) = [\![\varphi \wedge \neg\psi]\!]$. Since $t \cap \mathsf{f}$ is an argument we know that the agent has an argument for $\varphi \wedge \neg\psi$,

$$\mathcal{M}, w \vDash \boxdot(\varphi \wedge \neg\psi) \ .$$

□

2.5.5. Proposition.

$$\vDash \mathcal{T}\varphi \to \forall(\Box\varphi \to \mathcal{T}\varphi) \ . \tag{2.3}$$

Proof:
The proof of the validity of formula 2.3 follows once the following two facts are observed:

- the satisfaction of $\mathcal{T}\varphi$ in some possible world implies that there is an argument $\mathsf{f} \in \mathsf{LFP}_\tau$ such that $\mathsf{f} \subseteq [\![\varphi]\!]$;
- given any possible world w, if t is a factive argument with respect to w (which is implied by the satisfaction of $\Box\varphi$ in w), then $t \cup \mathsf{f}$ is also a factive argument with respect to w belonging to LFP_τ.

The membership of $t \cup \mathsf{f}$ in LFP_τ relies on the fact that $\mathsf{f} \in \mathsf{LFP}_\tau$ and Proposition 2.3.8 in Section 2.3.3. □

In Section 2.3, we studied the relationship between the justified belief *Bel* in Baltag, Bezhanishvili, Özgün and Smets (2016*a*) and the grounded belief $\mathcal{B}$ (Proposition 2.3.11 and Corollary 2.3.13). We have mentioned that justified belief *Bel* is definable in language $\mathcal{L}_{\Box\forall}$ ($Bel\,\varphi := \forall \Diamond \Box\varphi$). So Corollary 2.3.13 can be expressed in ABL by the following valid formula:

$$\vDash \forall \Diamond \Box\varphi \to \exists \mathcal{T}\varphi \ . \tag{2.4}$$

Except for the axioms we have discussed, the other axioms and rules in ABS are self-explanatory. More interesting theorems and rules can be derived from the axiom system ABS, as the formula in Proposition 2.5.3 and the formula $\forall(\varphi \to \psi) \to (T\varphi \to T\psi)$. We prove the last one to illustrate how the system works. More complicated theorems and their derivations in the ABS can be found in Section 3.5.

2.5.6. Fact.

$$\vdash_{\mathrm{ABS}} \forall(\varphi \to \psi) \to (\mathcal{T}\varphi \to \mathcal{T}\psi)$$

Proof:

(1) $\vdash (\mathcal{T}\varphi \wedge \forall(\varphi \to \psi)) \to \mathcal{T}(\varphi \wedge \forall(\varphi \to \psi))$	axiom $(\mathcal{T}\varphi \wedge \forall\psi) \to \mathcal{T}(\varphi \wedge \forall\psi)$
(2) $\vdash (\varphi \wedge \forall(\varphi \to \psi)) \to \psi$	axiom T for $\forall$ and Modus Ponens
(3) $\vdash \mathcal{T}(\varphi \wedge \forall(\varphi \to \psi)) \to \mathcal{T}\psi$	(2) and from $\varphi \to \psi$ infer $\mathcal{T}\varphi \to \mathcal{T}\psi$
(4) $\vdash \forall(\varphi \to \psi) \to (T\varphi \to T\psi)$	(1),(3) and Modus Ponens

□

The axiom system ABS is strongly complete and sound with respect to the class of topological argumentation models.

2.5.7. Theorem. *For all $\varphi \in \mathcal{L}_{\forall\mathcal{T}\Box}$ and all $\Phi \subseteq \mathcal{L}_{\forall\mathcal{T}\Box}$,*

$$\Phi \vdash_{\text{ABS}} \varphi \qquad \textit{if and only if} \qquad \Phi \vDash \varphi$$

Due to the similarity between ABL and the logics with respect to the topological semantics studied in (Baltag, Bezhanishvili, Özgün and Smets, 2016*a*), we will prove the completeness result with respect to the class of quasi-a-models, analogous to the quasi-model of Özgün (2017, Section 5.6.5). The differences, nevertheless, are also noticeable. For example, the attack relation in the quasi-a-model is nowhere to be found in the quasi-model. And the quasi-a-model we construct in the proof is *Alexandroff* (we will define it below), while the quasi-model is not. So we do not need to adjust the quasi-a-model constructed in the proof to make it *Alexandroff*, which will give us the modal equivalence between the quasi-a-model and the topological argumentation model with respect to the language $\mathcal{L}_{\forall\mathcal{T}\Box}$.

2.5.8. Definition (Alexandroff space). A topological space (X, τ) is an *Alexandroff space* if τ is closed under arbitrary intersections, i.e., $\bigcap A \in \tau$ for all $A \subseteq \tau$.

2.5.9. Definition (Quasi-a-model). A *quasi-a-model* is a tuple

$$\mathsf{M} = (W, E_0, \leqslant, \twoheadleftarrow, V)$$

where $(W, E_0, \tau_{E_0}, \twoheadleftarrow, V)$ is a topological argumentation model and $\leqslant$ is a preorder such that for every $e \in E_0$, if $u \in e$, then $v \in e$ for all $v \in W$ satisfying $u \leqslant v$.

The semantics of language $\mathcal{L}_{\forall\mathcal{T}\Box}$ in the quasi-a-model stays the same except for $\Box$:

$$\mathsf{M}, w \vDash \Box\varphi \quad \text{iff} \quad \text{for all } v \in W \text{ such that } w \leqslant v, \mathsf{M}, v \vDash \varphi \ .$$

The detailed completeness proof can be found in Section A.2. Here we introduce some definitions and results which will play a key role not only in the proof of the completeness result but also in Chapter 3.

2.5.10. Definition (Specification preorder). Given a topological space (X, τ), its *specification preorder* is defined by

$$x \sqsubseteq_\tau y \text{ iff for all } t \in \tau, x \in t \text{ implies } y \in t$$

where $x, y \in X$.

2.5.11. Definition (Alexandroff quasi-a-model). A quasi-a-model $\mathsf{M} = (W, E_0, \leqslant, \twoheadleftarrow, V)$ is called *Alexandroff* if the topological space in it (W, τ_{E_0}) is *Alexandroff* space and $\leqslant = \sqsubseteq_\tau$.

2.5.12. Proposition. *Given an Alexandroff quasi-a-model* $\mathsf{M} = (W, E_0, \leqslant, \twoheadleftarrow, V)$*, let* $\mathcal{M} = (W, E_0, \tau_{E_0}, \twoheadleftarrow, V)$ *be its corresponding topological argumentation model, then for all* $\varphi \in \mathcal{L}_{\forall \mathcal{T} \Box}$,

$$[\![\varphi]\!]_{\mathsf{M}} = [\![\varphi]\!]_{\mathcal{M}} \quad .$$

Proof:
The proof for this proposition is the same to the proof of Proposition 5.6.14 in Özgün (2017), because the only new operator $\mathcal{T}$ has the same truth condition in both M and $\mathcal{M}$. $\Box$

2.6 Failure of Closure and Rational Belief

It has been shown in Section 2.3.3 that grounded belief fails to satisfy the principle of closure under conjunction (also called the *conjunction rule*), which is counted by many (Hintikka, 1962; Levi, 1973) as a fundamental rationality postulate. In this section, we examine the principle and its failure for grounded belief in more detail.

Usually, the debate about the conjunction rule for categorical/binary/all-or-nothing belief is raised when people try to bridge all-or-nothing belief with the agent's numerical belief/degree of belief.[9] The lottery paradox (Kyburg, 1961) reveals the tension between categorical belief and numerical belief, which forces us to choose either giving up the conjunction rule or allowing belief of contradiction.

Grounded belief, nevertheless, is defined in a purely qualitative way, which has no explicit dependence on any numerical measurement. What are the causes and consequences of the failure of conjunction rule for grounded belief?

It is clear that the failure of conjunction rule for grounded belief is not caused by the agent's lack of logical reasoning ability, because we assume that the agent

[9] An example is the Lockean thesis (cf. Foley (2009); also, see Shi, Smets and Velázquez-Quesada (2017*a*, Section 5) for a brief comparison between grounded belief and belief based on the Lockean thesis).

is logically omniscient and the topology in the topological argumentation model represents clearly how the agent can perfectly combine her evidence logically. Moreover, the topological argumentation model explicitly represents the distinction and connection between the agent's logical reasoning, argumentational reasoning and grounded belief, which unloads the burden of deductive reasoning from the agent's belief. Although the agent's belief relies and thus reflects the agent's reasoning ability, it does not determine it. So the failure of conjunction rule for grounded belief should not imply the agent's failure of doing deductive reasoning.

What is then the real severe consequence of failing to be closed under conjunction? As we have indicated at the end of Section 2.3, it is the inconsistency. Grounded belief follows the *mutual consistency principle*:

> If S is a body of reasonably accepted statements, then there is no pair of members of S, s_1 and s_2, such that every statement of the language follows from s_1, s_2 as premises.

but not the *strong principle of consistency*:

> If S is a body of reasonably accepted statements, then there is no finite subset of S, $s_1, \ldots, s_n$, such that every statement of the language follows from $s_1, \ldots, s_n$ as premises. (Kyburg, 1970, p.59)

There seems to be a dilemma embedded in the agent's choices of her belief – either keep following the strong principle of consistency but stay conservative about what to believe (for example, justified belief) or accept more statements as beliefs but embrace some inconsistency in the body of beliefs (for example, grounded belief).

Should we rest easy and be satisfied with a notion of belief with mutual consistency and failure of conjunction rule? A lot of discussions has been made on this question (Kyburg, 1970; Leitgeb, 2017). While staying neutral to the above question, we can ask alternatively: could there be some other notions of belief striking a better balance between the agent's boldness and the consistency of her beliefs than grounded belief? We pursue such a notion in the next Chapter.

2.7 Conclusion

In this section, we proposed a notion of belief – grounded belief, which shows how the agent starts with her 1-consistent basic evidence collection and reaches the 2-consistent/mutually consistent collection of successfully defended pieces of evidence by argumentational reasoning.

We have given a complete axiomatic characterisation of grounded belief. Moreover, we axiomatized the logic in which both grounded belief and arguments can be expressed. The interrelationship between grounded belief and argument is reflected in the axiom system.

One of the reasons for introducing the attack relation is to weigh mutually contradictory pieces of evidence differently. The same thing can also be done by introducing a preference relation between pieces of evidence, as suggested by van Benthem and Pacuit (2011). The difference is that a preference relation can exist between two pieces of evidence even if they do not contradict each other, while the attack relation in this chapter is required by us to exist only between mutually contradictory pieces of evidence. The way we define grounded belief based on the attack relation may provide new insights on how the work in van Benthem and Pacuit (2011) can be further developed by adding a preference relation.

Another direction for a further study is the dynamics of argumentation, which includes not only the dynamics of arguments but also the dynamics of attack relation between arguments. About the dynamics of arguments, van Benthem and Pacuit (2011) has already investigated several different possible operators through the approach of dynamic epistemic logic as we briefly mentioned in Section 2.2. The addition of attack relation certainly brings more interplay between different arguments and thus a more complex relationship between arguments and belief. Some papers in formal argumentation theory (Liao, Jin and Koons, 2011; Liao, 2013; Baroni, Giacomin and Liao, 2014) have studied the dynamics of argumentation from a computational perspective. Together with the works on the dynamics of belief and evidence (Van Ditmarsch, van Der Hoek and Kooi, 2008; van Benthem, 2011; van Benthem and Pacuit, 2011; Baltag, Bezhanishvili, Özgün and Smets, 2016*b*), they lay the ground for further exploration of the dynamics of argumentation and belief in the topological setting.

Chapter 3

Belief by Default and Evidence

A group of blind men heard that a strange animal, called an elephant, had been brought to the town, but none of them were aware of its shape and form. Out of curiosity, they said:"We must inspect and know it by touch, of which we are capable". So, they sought it out, and when they found it they groped about it.

In the case of the first person, whose hand landed on the trunk, said "This being is like a thick snake".

For another one whose hand reached its ear, it seemed like a kind of fan.

Another person, whose hand was upon its leg, said, the elephant is a pillar like a tree-trunk.

The blind man who placed his hand upon its side said, "elephant is a wall". Another who felt its tail described it as a rope.

The last felt its tusk, stating the elephant is that which is hard, smooth and like a spear.

(*Parable:Blind men and an elephant*)[1]

[1] Wikipedia (2017)

3.1 Introduction

In the previous chapter, we take argumentational reasoning as a way of resolving the conflicts between different pieces of evidence. By putting priority on one of the two arguments which attack each other, rather than weighing them equally, argumentational reasoning prevents the agent from being too unopinionated and too cautious about what to believe. In comparison with the notion of justified belief as proposed in Baltag, Bezhanishvili, Özgün and Smets (2016*a*), agents will have more grounded beliefs, but on the side of consistency, they have to sacrifice a bit. Recall that grounded belief is only 2-consistent, but fails to be 3-consistent, as illustrated in Example 2.3.2,

In this chapter, we try to resolve the tension by taking into consideration a significant piece in the puzzle of human reasoning – default reasoning.

As human beings, we reason by following some inference rules. For example, from A and B infer A. But mostly the inference rules we follow are not universally valid. Consider for example the two statements, birds can fly and mammals do not lay eggs. We know that not all birds can fly, and some mammals do lay eggs. However, we still accept these inference rules in the sense of applying them to our evidence and getting to the conclusions when there is no available information in conflict with the conclusions. These rules are called default rules and their use for getting further information and making judgements is called default reasoning.

The blind men in the parable "blind men and an elephant" are doing default reasoning in the sense of reaching their conclusions based on limited evidence. The reason for them to be the objects of ridicule is not the fact that they jump to the conclusion, but rather the fact that they jump to the conclusion from a wrong premise – what they touched is the whole elephant. So the critical question when it comes to default reasoning is not *whether* it should be used but rather *how* it should be used.

To be more specific, we ask in this chapter, how default reasoning should be coordinated with the agent's argumentational reasoning so that the agent's belief based on them is not as stringent as justified belief while maintaining its full consistency.

The lesson we can learn from the blind men in the parable is that premises should play as important a role as default rules in default reasoning. Proper default reasoning depends on the reliability of its premises. Which part of the agent's evidence can serve as premises? Recall the function of the agent's argumentational reasoning. It fits very well the task of picking pieces of evidence eligible as premises of default reasoning. Then, should we take what follows by default reasoning from those premises whose quality is controlled by argumentational reasoning as the agent's belief? What is the relationship between the agent's default reasoning and her belief, given that argumentational reasoning filters out those pieces of evidence unqualified for premises?

The proposal in this chapter suggests that the agent's beliefs should be supported by *all* qualified pieces of evidence in the sense of following from them by default reasoning. For example, if the blind men in the parable were asked to touch all other parts of the elephant, I guess none of them would stick to their original answers because their original answers cannot be supported by all qualified pieces of evidence which were available to them.

Outline of Chapter 3 There are different ways of representing default rules and formalizing default reasoning in the literature (Geffner, 1992; Boutilier, 1994*a*; Veltman, 1996). We will follow the "conditional approach to default reasoning" proposed in Boutilier (1994*a*), which can be traced back to Stalnaker (1968); Lewis (1973); Burgess (1981); Veltman (1985).

Section 3.2 briefly introduces the formalism in Boutilier (1994*a*), where default rules are represented as conditionals (if Tweety is a bird, then *normally* Tweety can fly) and semantically interpreted through a normality order on possible worlds. Section 3.3 and Section 3.4 then address the following two questions respectively:

- how does the formalism of default reasoning fit into the topological setting?
- how does default reasoning apply to the agent's evidence and, together with logical reasoning and argumentational reasoning, contribute to the agent's beliefs?

It will be shown in Section 3.5 that the new notion of belief – full-support belief – resulted from taking default reasoning into account in the topological setting satisfies the conjunction rule and thus the strong principle of consistency (page 37). Moreover, agents can have more full-support beliefs than justified beliefs, while having more grounded beliefs than full-support beliefs. All the results about the logical properties of full-support belief and its relationship with justified belief and grounded belief can be achieved by an application of the logic of argument and belief (ABL) developed in Section 2.5.

3.2 Conditional Approach to Default Reasoning

Following Boutilier (1994*a*), we represent default rules by conditionals using a connective $\rightsquigarrow$ in the objective language, in contrast with its representation as a consequence relation in Kraus, Lehmann and Magidor (1990). For example, "if Tweety is a bird, then normally Tweety can fly" can be represented as $b \rightsquigarrow f$. To express this form of default rules, we first focus on the following formal language.

3.2.1. Definition. Given a set of atomic propositions At, the language of default rules $\mathcal{L}_{\rightsquigarrow}$ is given by the following grammar:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \varphi \rightsquigarrow \varphi$$

where $p \in \mathsf{At}$.

The semantic model is given by imposing a reflexive and transitive order $\leqslant$ on a set of possible worlds W. $v \leqslant u$ means that possible world u is at least as normal as possible world v.[2] $v \leqslant u$ and $u \nleqslant v$ will be abbreviated to $v < u$ (u is more normal than v), and $v = u$ (v and u are equally normal) is the abbreviation for $v \leqslant u$ and $u \leqslant v$.

3.2.2. Definition (Normality Model (Boutilier, 1994*a*)). A *normality model* $\mathcal{M}$ is a triple $(W, \leqslant, V)$ where W is a non-empty set of possible worlds. The normality order $\leqslant \subseteq W \times W$ is a reflexive and transitive order on W, and $V : \mathsf{At} \to 2^W$ is a valuation function assigning to each atomic proposition a subset of W.

The truth conditions for the propositional fragment of $\mathcal{L}_{\rightsquigarrow}$ are defined in a standard way. For $\varphi \rightsquigarrow \psi$, roughly speaking, given a normality model, it can be semantically evaluated in the model by checking whether those *most normal* $[\![\varphi]\!]$-worlds are also $[\![\psi]\!]$-worlds.[3] To make it more precise, we define the set of most normal P-worlds in a normality model $\mathcal{M}$ as follows [4]

$$\max_{\leqslant}(P) = \{v \in P \subseteq W \mid \text{ there is no } u \in P \text{ such that } v < u\} \ .$$

One problem of the definition given above for $\varphi \rightsquigarrow \psi$ is that $\max_{\leqslant}(P)$ can be empty even if P is not empty, which will make $\varphi \rightsquigarrow \bot$ true where $[\![\varphi]\!] = P$. There are two ways of avoiding this problem. The first one is straightforward – requiring that for all nonempty subsets P of W $\max_{\leqslant}(P) \neq \varnothing$. This assumption is called "limit assumption" in the literature. The second way is generalising the truth condition of $\varphi \rightsquigarrow \psi$ so that it functions well even when $\max_{\leqslant}[\![\varphi]\!]$ is empty.

3.2.3. Definition (Truth condition of $\rightsquigarrow$ (Boutilier, 1994*a*)). Given a normality model, $P \rightsquigarrow Q$ holds if and only if for all P-worlds v in the model there is a P-world u such that $v \leqslant u$ and for all P-worlds x satisfying $u \leqslant x$, $x \in Q$.

If the normality model satisfies the limit assumption then Definition 3.2.3 is equivalent to the truth condition of $\rightsquigarrow$ using $\max_{\leqslant}$. The difference occurs when $\max_{\leqslant}(P)$ is empty for some $P \subseteq W$. The emptiness of the set $\max_{\leqslant}(P)$ implies that for all $v \in P$, there is an infinite ascending chain $v \leqslant v_1 \leqslant v_2 \leqslant \ldots$. Definition 3.2.3 only checks whether in the model for every ascending chain of this kind there is a world in it, from which on all the worlds are Q-worlds.

---

[2] Note that we do not follow the convention in the literature of conditional logic where v is taken to be at least as normal as u if $v \leqslant u$.

[3] We will write P-worlds to denote those worlds belonging to $P \subseteq W$.

[4] This way of defining a conditional connective in an order structure can be found in the literature on conditionals, for example, Stalnaker (1968); Lewis (1973). In these works on conditionals, conditionals are usually defined by a selection function, which is a more general way. Similar semantics can also be found in several works on nonmonotonic reasoning, for example, Shoham (1987); Kraus, Lehmann and Magidor (1990). For a more complete review of the relevant literature, we refer readers to Boutilier (1994*b*, Section 1.1).

A further observation, which has appeared in Boutilier (1994*a*); Baltag and Smets (2008) and so on, is that Definition 3.2.3 can be expressed in the following language.

3.2.4. Definition. Given a set of atomic propositions At, the language $\mathcal{L}_{[\leqslant]\forall}$ is generated from the following grammar:

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid [\leqslant]\varphi \mid \forall\varphi$$

where $p \in \mathsf{At}$.

In this language $[\leqslant]$ is the standard normal modality defined by the binary relation $\leqslant$ and $\forall$ is a universal modality which in this single-agent case can be taken as the S5 knowledge operator.

3.2.5. Definition. Given a normality model $\mathcal{M}$, and a possible world w in it,

$$\begin{array}{lll} \mathcal{M}, w \vDash [\leqslant]\varphi & \text{iff} & \text{for all } v \in W \text{ such that } w \leqslant v,\ v \in [\![\varphi]\!]; \\ \mathcal{M}, w \vDash \forall\varphi & \text{iff} & W \subseteq [\![\varphi]\!]. \end{array}$$

Then $\rightsquigarrow$ can be syntactically defined in $\mathcal{L}_{\square\forall}$ as follows:

$$\varphi \rightsquigarrow \psi := \forall(\varphi \to \langle\leqslant\rangle(\varphi \wedge [\leqslant](\varphi \to \psi))) \tag{3.1}$$

where $\langle\leqslant\rangle := \neg[\leqslant]\neg$. Formula 3.1 expresses the same meaning as Definition 3.2.3.

Given a normality model $\mathcal{M}$ and a formula $\varphi \in \mathcal{L}_{[\leqslant]\forall}$ as a description of the agent's evidence, it can be decided what follows from the agent's default reasoning and her evidence, that is, the set $\{\psi \in \mathcal{L}_{[\leqslant]\forall} \mid \mathcal{M} \vDash_{[\leqslant]\forall} \varphi \rightsquigarrow \psi\}$. Therefore, the agent's default reasoning can be formalised very elegantly in this framework. The problem now is to fit this framework into the topological setting where the fuel of the agent's reasoning – evidence – is explicitly modelled. After all, without any evidence, there is nowhere to apply default reasoning, and more seriously, there may be nowhere to learn any default rules.

3.3 Default Rules, Evidence and Knowledge

Section 3.2 tells us that the default rules accepted by the agent are encoded in a normality order on a set of possible worlds. How should we consider the agent's normality order in a topological argumentation model (Definition 2.3.1)?

The topological argumentation model characterises the agent's evidence more explicitly than the normality model. So when considering how the agent ranks her epistemically possible worlds according to their normality in a topological argumentation model, we have to take into account the relationship between the agent's evidence and her normality ranking. At least, the normality order should respect the agent's evidence.

Given a topological argumentation model $\mathcal{M} = (W, E_0, \tau_{E_0}, \twoheadleftarrow, V)$, take arbitrary two possible worlds u and v, if all arguments in τ against u [5] are also against v, then u should be at least as plausible as v to the agent (van Benthem and Pacuit, 2011).[6] This way of ranking possible worlds is exactly how the specification preorder (Definition 2.5.10) for a topological space is defined:

$$v \sqsubseteq_\tau u \text{ iff for all } t \in \tau, u \notin t \text{ implies } v \notin t \ .$$

Note that this order is reflexive and transitive but not necessarily connected (for all $v, u \in W$, either $v \sqsubseteq_\tau u$ or $u \sqsubseteq_\tau v$).

Let $v \sqsubseteq_\tau u$ denote that u is at least as plausible as v. We will abbreviate $v \sqsubseteq_\tau u$ and $u \not\sqsubseteq_\tau v$ to $v \sqsubset_\tau u$ and abbreviate $v \sqsubseteq_\tau u$ and $u \sqsubseteq_\tau v$ to $v \sqsupseteq\!\!\sqsubseteq_\tau u$. Then what is the relationship between normality and plausibility?

One possible answer is that the normality order should not go against the plausibility order. To elaborate, if u is at least as normal as v then u is not less plausible than v, and if u is incomparable with v with respect to normality then u is incomparable with v with respect to plausibility:

$$\begin{gathered} v \leqslant u \text{ implies } u \not\sqsubset_\tau v \text{ and} \\ v \nleqslant u \text{ and } u \nleqslant v \text{ imply } v \not\sqsubseteq_\tau u \text{ and } u \not\sqsubseteq_\tau v \end{gathered} \tag{3.2}$$

Or we can require further that the normality order does not go beyond the plausibility, that is, the other way around of (3.2):

$$\begin{gathered} v \sqsubseteq_\tau u \text{ implies } u \nless v \text{ and} \\ v \not\sqsubseteq_\tau u \text{ and } u \not\sqsubseteq_\tau v \text{ imply } v \nleqslant u \text{ and } u \nleqslant v \end{gathered} \tag{3.3}$$

(3.2) and (3.3) together imply that

$$v \leqslant u \text{ if and only if } v \sqsubseteq_\tau u \ . \tag{3.4}$$

Since the agent's normality order should respect her evidence, we take (3.2) to be a reasonable condition on the relationship between the agent's normality order and her plausibility order. How about (3.3)? It is fully reasonable to think that the normality order is finer than the plausibility order because of some information which cannot be conveyed in the topological representation of evidence. However, to illustrate how default reasoning can be applied to the agent's evidence, it may be wise to take both condition 3.2 and condition 3.3

[5] For all $t \in \tau$, t is against u if and only if $u \notin t$.

[6] The plausibility order is widely used in modal logic as a way of modeling conditional belief (van Benthem, 2007; Baltag and Smets, 2008). The idea of inducing a plausibility order from a familiy of evidence sets is originally proposed in van Benthem and Pacuit (2011).

(hence condition 3.4) which will make the formalization simpler rather than to pursue a more general representation of normality orders. However, we note that taking $\leqslant = \sqsubseteq_\tau$ does not implicate that plausibility and normality should be conceptually identical.

Recall that in Section 2.5, we introduce the logic of argument and belief and prove it is sound and strongly complete not only with respect to the class of topological argumentation models but also with respect to the class of Alexandroff quasi-a-models (Definition 2.5.11). In an Alexandroff quasi-a-model, we have an order $\leqslant$, which is required to be equal to $\sqsubseteq_\tau$, and the topological space in the model is required to be an Alexandroff space (Definition 2.5.8).

3.3.1. FACT. Given an Alexandroff quasi-a-model and any world w in it, since the topological space in it is an Alexandroff space, it follows by Özgün (2017, Proposition 3.1.4.2) that

$$\text{for all } v \in W \text{ such that } w \sqsubseteq_\tau v,\ v \in P \textbf{ iff } \text{there is } t \in \tau_{E_0} \text{ such that } w \in t \subseteq P.$$

This means that we can express default rules in the logic of argument and belief with respect to Alexandroff quasi-a-models.

3.3.2. FACT. Given an Alexandroff quasi-a-model, if we take $\leqslant$ in it as the agent's normality order,

$$\varphi \rightsquigarrow \psi := \forall(\varphi \to \Diamond(\varphi \wedge \Box(\varphi \to \psi)))$$

where $\Box$ is an operator for *factive* evidence and $\forall$ is an operator for the agent's knowledge in the logic of argument and belief.

Therefore, in the case where the agent's normality order coincides with her plausibility order extracted from her topology of evidence and the topological space representing her evidence is an Alexandroff space, the relationship between the agent's default rules, knowledge and evidence is crystallised in the above formula.

By integrating a normality order into the topological argumentation model, according to the way in which default reasoning is modelled, what the agent can conclude stay fully consistent when applying her default rules to each piece of evidence separately. The problem appears when we put the agent's different pieces of evidence together. Some of these pieces conflict with each other, which imply that the conclusions follow from them by default reasoning would mostly also conflict.

We manage to resolve the inconsistency in the agent's evidence partly by picking out those pieces of evidence which can be defended successfully; i.e. those in the least fixed point of a characteristic function d (Definition 2.3.4). This set of pieces of evidence are at least pairwise consistent. However, a direct application of default rules on successfully defended evidence does not help resolve the inconsistency fully. Example 2.3.14 illustrates such a situation.

3.3.3. Example (Example 2.3.14 continued). In the example, all possible worlds are incomparable with each other with respect to $\sqsubseteq_\tau$. And the least fixed point is $\{\{1,2\},\{2,3\},\{3,1\},\{1,2,3\}\}$. So according to the definition of $\rightsquigarrow$ (Definition 3.2.3),

$$\{1,2\} \rightsquigarrow \{1,2\} \quad \{2,3\} \rightsquigarrow \{2,3\} \quad \{3,1\} \rightsquigarrow \{3,1\} \ .$$

But $\{1,2\} \cap \{2,3\} \cap \{3,1\} = \varnothing$.

In the next section, we will propose a way in which the application of the agent's default rules resolves the inconsistency among the agent's evidence and thus achieves a fully consistent set of conclusions. We take this set of conclusions as the agent's beliefs and show that the new notion of belief is not as conservative as justified belief.

3.4 Reason to Believe

To better illustrate the idea behind the new notion of belief featured in this section, we start with simple cases.

3.4.1 When the strongest argument in LFP_τ exists

Consider a topological argumentation model $\mathcal{M}$ whose LFP_τ includes an argument f such that for all $f' \in \mathsf{LFP}_\tau$, $f \subseteq f'$. In this case, how should the agent's default reasoning work (taking $\sqsubseteq_\tau$ as the normality order)? It seems reasonable to collect all the conclusions following from f by applying her default rules:

$$\{P \subseteq W \mid f \rightsquigarrow P \text{ holds in } \mathcal{M}\}$$

and take them as beliefs. After all, there is no argument in LFP_τ inconsistent with these conclusions. However, for some $f' \in \mathsf{LFP}_\tau$ with $f \subset f'$ there could be some $Q \in \{P \subseteq W \mid f \rightsquigarrow P \text{ holds in } \mathcal{M}\}$ such that $f' \rightsquigarrow Q$ does not hold in $\mathcal{M}$.

3.4.1. Example. For example, by deleting the attack relation from $\{1,3\}$ to $\{2\}$ in the topological argumentation model $\mathcal{M}_{123}$ of Example 2.3.2, we can get the argumentation framework illustrated in Figure 3.1.

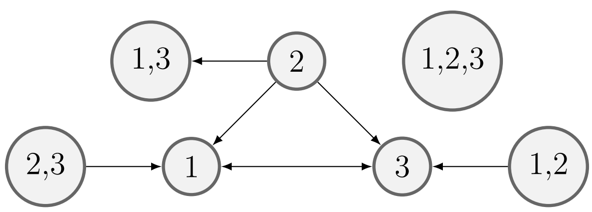

Figure 3.1: The agent's argumentation framework for $\mathcal{M}_{123}$ without $\{2\} \twoheadleftarrow \{1,3\}$

The new least fixed point

$$\mathsf{LFP}_\tau = \{\{2\},\{1,2\},\{2,3\},\{1,2,3\}\}$$

has $\{2\}$ as its strongest argument. Note that $\{1,2\} \rightsquigarrow \{2\}$ does not hold, because 1 and 2 are incomparable with respect to $\sqsubseteq_\tau$.

Should this affect the agents' acceptance of Q (in the example $Q = \{2\}$) as her belief? No, because the stronger premise f should have priority over f' when f and f' are both defended successfully. The priority of f as the premise in the agent's default reasoning over weaker argument f' is conditional on the membership of f in LFP_τ. The agent should exploit her successfully defended arguments as much as possible by default reasoning to extract more information. The agent's belief is thus extended by her default reasoning whose quality is controlled by the membership of the premises in LFP_τ. Thus we can summarise the agent's strategy of getting beliefs from arguments by default reasoning as follows.

> **Principle:** Strengthen the premise (coming from the agent's arguments) for default reasoning as much as possible while keeping it under successful defence.

The problem is that the structure of LFP_τ for topological argumentation models can be much more complicated than the case we have just considered where we can find the strongest argument in LFP_τ. What if there is no open $f \in \mathsf{LFP}_\tau$ such that $f \subseteq f'$ for all $f' \in \mathsf{LFP}_\tau$?

3.4.2 When there exists a set $\mathsf{Mini} \subseteq \mathsf{LFP}_\tau$

Consider a topological argumentation model $\mathcal{M}$ whose LFP_τ includes a set of arguments $\mathsf{Mini} \subseteq \mathsf{LFP}_\tau$ such that for all $f \in \mathsf{LFP}_\tau$ there is $m \in \mathsf{Mini}$ satisfying $m \subseteq f$ and for all $m \in \mathsf{Mini}$ there is no $f' \in \mathsf{LFP}_\tau$ satisfying $f' \subset m$. Note that in this subsection we assume that Mini exists.

Assume that Mini is not a singleton. For a more concrete example, we refer to Example 2.3.2 where $\mathsf{Mini} = \{\{1,2\},\{2,3\}\}$. To which argument in Mini should the agent apply her default rules? Observe that for all $m, m' \in \mathsf{Mini}$, if $m \rightsquigarrow Q$ and $m' \rightsquigarrow Q'$ both hold, then $Q \cap Q' \neq \varnothing$. So we may expect to define the agent's beliefs as those propositions which follow from one of the arguments in Mini:

> (a) the agent believes Q in the topological argument model $\mathcal{M}$ **if and only if** *there is* $m \in \mathsf{Mini}$ such that $m \rightsquigarrow Q$ holds in $\mathcal{M}$.

This way of defining the agent's belief, however, does not avoid the failure of the conjunction rule. To save the failure, we can strengthen the condition by requiring that Q follows from all arguments in Mini rather than only one of them:

(b) the agent believes Q in the topological argument model $\mathcal{M}$ **if and only if** *for all* $m \in \mathsf{Mini}$, $m \rightsquigarrow Q$ holds in $\mathcal{M}$.

The agent's belief in this definition is more robust in the sense of enduring all the tests – supported by all members in Mini through default reasoning. Moreover, it ensures that the agent's beliefs are closed under conjunction.

An interesting but a little shocking fact about the two ways of defining belief is stated in the following proposition.

3.4.2. PROPOSITION. *Given a topological argumentation model* $\mathcal{M}$*, take its specification order* $\sqsubseteq_\tau$ *as the normality order. If* $\mathsf{Mini} \subseteq \mathsf{LFP}_\tau$ *exists, then for all* $m \in \mathsf{Mini}$ *and* $P \subseteq W$*,* $m \subseteq P$ *if and only if* $m \rightsquigarrow P$ *holds in* $\mathcal{M}$.

This proposition says that the application of default rules to arguments in Mini does not extend what are supported by the arguments in Mini. It thus implies that the first way of defining belief in (a) is exactly the way we define grounded belief given the assumption that Mini exists. And the second way of defining belief in (b) amounts to taking the set $\{P \subseteq W \mid \bigcup_{m \in \mathsf{Mini}} m \subseteq P\}$ as the agent's beliefs.

The remaining task of this section is to remove the assumption of the existence of Mini. As we will see, a similar result (Theorem 3.4.6) to Proposition 3.4.2 also holds when Mini does not exist.

3.4.3 General case

3.4.3. EXAMPLE. For a concrete example where Mini does not exist, take $W = \mathbb{N} \cup \{\omega\}$ and let $E_0 = \{\varnothing\} \cup \{[n, \omega] \mid n \in \mathbb{N}\} \cup \{\{1\}\}$. The attack relation is specified as follows: $\{1\} \twoheadleftarrow [n, \omega]$ and $[n, \omega] \twoheadleftarrow \{1\}$ for all $n > 1$. The whole model [7] is illustrated in Figure 3.2, where most of the attack relations are left out for the simplicity of the graph. $\mathsf{LFP}_\tau = \{[n, \omega] \cup \{1\} \mid n \in \mathbb{N}\}$. For this model, Mini does not exist.

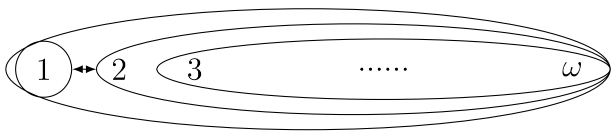

Figure 3.2: The model in Example 3.4.3

To deal with more general cases, we generalise the previous way of defining belief and propose the following formal definition.

[7] The attack relation is symmetric, so it can also be seen as a topological evidence model (Definition 2.2.6).

3.4.4. Definition (Full-support belief). Given a topological argumentation model $\mathcal{M}$, take $\sqsubseteq_\tau$ as the agent's normality order. The agent has *full-support belief* of $P \subseteq W$ if and only if *for all* arguments $f \in \mathsf{LFP}_\tau$, f can be strengthened to an argument f' belonging to LFP_τ such that $f' \rightsquigarrow P$ holds in $\mathcal{M}$.

Note that when Mini exists for the topological argumentation model, Definition 3.4.4 is equivalent to the condition that for all $m \in \mathsf{Mini}$, $f \rightsquigarrow P$ holds in the model. The equivalence hinges on the following fact.

3.4.5. Fact. Given a topological argumentation model $\mathcal{M}$, take $\sqsubseteq_\tau$ as the agent's normality order. For all $t, t' \in \tau$, if $t \subseteq t'$ and $t' \rightsquigarrow P$ holds in $\mathcal{M}$, then $t \rightsquigarrow P$ also holds in $\mathcal{M}$.

The following theorem is a generalization of Proposition 3.4.2.

3.4.6. Theorem. *Given a topological argumentation model $\mathcal{M}$, take $\sqsubseteq_\tau$ as the agent's normality order. If the topological space in $\mathcal{M}$ is an Alexandroff space, then for all arguments $f \in \mathsf{LFP}_\tau$, f can be strengthened to an argument f' belonging to LFP_τ such that*

$$f' \rightsquigarrow P \text{ holds in } \mathcal{M}$$

if and only if *for all arguments $f \in \mathsf{LFP}_\tau$, f can be strengthened to an argument f'' belonging to LFP_τ such that*

$$f'' \subseteq P \ .$$

We will provide two ways of proving this theorem. One is a semantic proof which is given in Appendix B, and the other is a syntactic proof which makes use of axiom system ABS for the logic of argument and belief (see Section 2.5). The use of syntactic proof relies on expressing conditions in the theorem in the language $\mathcal{L}_{\forall\mathcal{T}\Box}$. So in the next section, we show how this theorem can be expressed and proved in the logic of argument and belief (ABL). Moreover, we prove the logical relation between justified belief, grounded belief and full-support belief.

3.5 Full-support Belief in ABL

In this section, we show that full-support belief can be formulated in the logic of argument and belief with respect to the class of Alexandroff quasi-a-models. Moreover, since axiom system ABS is sound and complete with respect to the class of Alexandroff quasi-a-models, we are allowed to make use of ABS to prove Theorem 3.4.6. The syntactic analysis fully elucidates how justified belief, grounded belief and full-support belief relate to each other logically.

In this section, we will always work with the class of Alexandroff quasi-a-models (abbreviated to AQA models), which takes $\sqsubseteq_\tau$ as the agent's normality order and requires that the topological space in the model is an Alexandroff space.

In the class of AQA models "the agent has full-support belief of φ" can be expressed in $\mathcal{L}_{\forall\mathcal{T}\Box}$ by $\forall\widehat{\mathcal{T}}\mathcal{T}\Diamond\Box\varphi$.

3.5.1. PROPOSITION. *Let $\mathcal{B}_f$ be an operator for full-support belief.*

$$\vDash_{AQA} \mathcal{B}_f\,\varphi \leftrightarrow \forall \widehat{\mathcal{T}}\,\mathcal{T} \Diamond \Box\,\varphi$$

where $\widehat{\mathcal{T}} \coloneqq \neg\,\mathcal{T}\,\neg$.

Proof:
We only outline the main intuition behind the proof. The details can be easily filled out.

Recall the definition of full-support belief:

- for all $f \in \mathsf{LFP}_\tau$, f can be strengthened to an argument f' belonging to LFP_τ such that
- $f' \rightsquigarrow [\![\varphi]\!]$.

First notice that "for all $f \in \mathsf{LFP}_\tau$, f can be strengthened to an argument f' belonging to LFP_τ such that $f' \subseteq [\![\varphi]\!]$" can be expressed by $\forall \widehat{\mathcal{T}}\,\mathcal{T}\,\varphi$ To see why this is the case, recall the reason why justified belief in Baltag, Bezhanishvili, Özgün and Smets (2016*a*), defined as "for every argument $t \in \tau$, t can be strengthened to an argument $t' \in \tau$ such that $t \subseteq [\![\varphi]\!]$, can be expressed by $\forall \Diamond \Box\varphi$.

Second, notice that "for all $t \in \mathsf{LFP}_\tau$, $t \rightsquigarrow [\![\varphi]\!]$ holds in $\mathcal{M}$" is equivalent to

$$\mathcal{M} \vDash_{AQA} \forall\,\mathcal{T} \Diamond \Box\,\varphi \ .$$

Statement "for all arguments $t \in \mathsf{LFP}_\tau$, $t \rightsquigarrow [\![\psi]\!]$ holds in $\mathcal{M}$" semantically means that in $\mathcal{M}$ for all arguments $t \in \mathsf{LFP}_\tau$, for all $v \in t$, there is $u \in t$ such that $v \sqsubseteq_\tau u$ and for all $w \in t$ satisfying $u \sqsubseteq_\tau w$, $w \in [\![\psi]\!]$. So it is equivalent to $\mathcal{M} \vDash_{AQA} \forall\,\mathcal{T} \Diamond \Box\,\psi$.

Combining these two observations, it follows that $\vDash_{AQA} \mathcal{B}_f\,\varphi \leftrightarrow \forall \widehat{\mathcal{T}}\,\mathcal{T} \Diamond\Box\varphi$. $\Box$

We then reformulate Theorem 3.4.6 as follows:

$$\vDash_{AQA} \forall \widehat{\mathcal{T}}\,\mathcal{T} \Diamond \Box\,\varphi \leftrightarrow \forall \widehat{\mathcal{T}}\,\mathcal{T}\,\varphi \ . \tag{3.5}$$

Next we derive $\forall \widehat{\mathcal{T}}\,\mathcal{T} \Diamond \Box\,\varphi \leftrightarrow \forall \widehat{\mathcal{T}}\,\mathcal{T}\,\varphi$ from axiom system ABS.

First we prove

3.5.2. THEOREM.

$$\vdash_{ABS} \exists\,\mathcal{T} \Diamond \Box\,\varphi \rightarrow \exists\,\mathcal{T}\,\varphi$$

Proof:
The key is the axiom (we have unfolded $\mathcal{B}$ to $\exists\,\mathcal{T}$ and $\boxdot$ to $\exists\Box$)

$$\exists\,\mathcal{T}\,\varphi \wedge \neg\exists\,\mathcal{T}\,\psi \wedge \forall((\varphi \wedge \psi) \rightarrow \Box(\varphi \wedge \psi)) \rightarrow \exists\,\Box\,(\varphi \wedge \neg\psi) \ .$$

The first step is to prove that

$$\vdash \forall((\Box \Diamond \Box\varphi \wedge \Box\varphi) \rightarrow \Box(\Box \Diamond \Box\varphi \wedge \Box\varphi)) \ .$$

We will use CPL to indicate the use of tautologies in classical propositional logic.

(1) $\Box\Diamond\Box\varphi \to \Box\Box\Diamond\Box\varphi$	Axiom 4 for $\Box$
(2) $\Box\varphi \to \Box\Box\varphi$	Axiom 4 for $\Box$
(3) $(\Box\Diamond\Box\varphi \wedge \Box\varphi) \to \Box(\Box\Diamond\Box\varphi \wedge \Box\varphi)$	(1),(2), axiom K for $\Box$ and CPL
(4) $\forall((\Box\Diamond\Box\varphi \wedge \Box\varphi) \to \Box(\Box\Diamond\Box\varphi \wedge \Box\varphi))$	Necessitation rule for $\forall$

The second step is to prove that

$$\vdash \neg\exists\Box(\Box\Diamond\Box\varphi \wedge \neg\Box\varphi)$$

(1) $\Box\Box\Diamond\Box\varphi \to \neg\Box\neg\Box\varphi$	Axiom T for $\Box$ and $\Diamond := \neg\Box\neg$
(2) $\neg(\Box\Box\Diamond\Box\varphi \wedge \Box\neg\Box\varphi)$	(1) and CPL
(3) $(\Box\Box\Diamond\Box\varphi \wedge \Box\neg\Box\varphi) \leftrightarrow \Box(\Box\Diamond\Box\varphi \wedge \neg\Box\varphi)$	Axiom K for $\Box$
(4) $\neg\Box(\Box\Diamond\Box\varphi \wedge \neg\Box\varphi)$	(2),(3) and CPL
(5) $\forall\neg\Box(\Box\Diamond\Box\varphi \wedge \neg\Box\varphi)$	(4) and Necessitation rule for $\forall$
(6) $\neg\exists\Box(\Box\Diamond\Box\varphi \wedge \neg\Box\varphi)$	(5), $\exists := \neg\forall\neg$ and CPL

The third step is to apply the key axiom we mentioned to the theorems achieved in the previous steps to get

$$\vdash \exists\mathcal{T}\Box\Diamond\Box\varphi \to \exists\mathcal{T}\Box\varphi$$

by CPL

The fourth step is to prove that

$$\vdash \exists\mathcal{T}\Diamond\Box\varphi \to \exists\mathcal{T}\Box\Diamond\Box\varphi\ .$$

(1) $\mathcal{T}\Diamond\Box\varphi \to \mathcal{T}\mathcal{T}\Diamond\Box\varphi$	Axiom 4 for $\mathcal{T}$
(2) $\mathcal{T}\mathcal{T}\Diamond\Box\varphi \to \mathcal{T}\Box\Diamond\Box\varphi$	Axiom $\vdash \mathcal{T}\varphi \to \Box\varphi$ and Monotonicity for $\mathcal{T}$
(3) $\mathcal{T}\Diamond\Box\varphi \to \mathcal{T}\Box\Diamond\Box\varphi$	(1),(2) and CPL
(4) $\exists\mathcal{T}\Diamond\Box\varphi \to \exists\mathcal{T}\Box\Diamond\Box\varphi$	Axioms and rules for $\forall$

The last step is to conclude by what achieved in third and fourth step that

$$\vdash \exists\mathcal{T}\Diamond\Box\varphi \to \exists\mathcal{T}\varphi\ .$$

$\Box$

Second, we prove the direction from left to right of 3.5 (the other direction is relatively easy whose proof is thus omitted here).

3.5.3. Theorem.

$$\vdash_{ABS} \forall\widehat{\mathcal{T}}\mathcal{T}\Diamond\Box\varphi \to \forall\widehat{\mathcal{T}}\mathcal{T}\varphi$$

Proof:
The key is what we have just proved:

$$\vdash_{ABS} \exists\mathcal{T}\Diamond\Box\varphi \rightarrow \exists\mathcal{T}\varphi \tag{3.6}$$

The first step is to prove that

$$\vdash \exists\mathcal{T}\widehat{\mathcal{T}}\mathcal{T}\varphi \rightarrow \exists\mathcal{T}\mathcal{T}\varphi$$

which is equivalent to proving

$$\vdash \forall\widehat{\mathcal{T}}\widehat{\mathcal{T}}\neg\varphi \rightarrow \forall\widehat{\mathcal{T}}\mathcal{T}\widehat{\mathcal{T}}\neg\varphi$$

(1) $\forall\widehat{\mathcal{T}}\widehat{\mathcal{T}}\neg\varphi \rightarrow \forall\widehat{\mathcal{T}}\neg\varphi$	Axiom 4 for $\mathcal{T}$ and axioms for $\forall$
(2) $\forall\widehat{\mathcal{T}}\neg\varphi \rightarrow \forall\forall\widehat{\mathcal{T}}\neg\varphi$	Axiom 4 for $\forall$
(3) $\forall\forall\widehat{\mathcal{T}}\neg\varphi \rightarrow \forall\mathcal{T}\widehat{\mathcal{T}}\neg\varphi$	$\vdash \forall\varphi \rightarrow \mathcal{T}\varphi$
(4) $\forall\mathcal{T}\widehat{\mathcal{T}}\neg\varphi \rightarrow \forall\mathcal{T}\mathcal{T}\widehat{\mathcal{T}}\neg\varphi$	Axiom 4 for $\mathcal{T}$ and axioms for $\forall$
(5) $\forall\mathcal{T}\mathcal{T}\widehat{\mathcal{T}}\neg\varphi \rightarrow \forall\widehat{\mathcal{T}}\mathcal{T}\widehat{\mathcal{T}}\neg\varphi$	$\vdash \mathcal{T}\varphi \rightarrow \widehat{\mathcal{T}}\varphi$
(6) $\forall\widehat{\mathcal{T}}\widehat{\mathcal{T}}\neg\varphi \rightarrow \forall\widehat{\mathcal{T}}\mathcal{T}\widehat{\mathcal{T}}\neg\varphi$	(1)(2)(3)(4)(5)

The second step is to prove that

$$\vdash \forall\widehat{\mathcal{T}}\mathcal{T}\Diamond\Box\varphi \rightarrow \exists\mathcal{T}\Diamond\Box\varphi \ .$$

(1) $\forall\widehat{\mathcal{T}}\mathcal{T}\Diamond\Box\varphi \rightarrow \forall\mathcal{T}\widehat{\mathcal{T}}\mathcal{T}\Diamond\Box\varphi$	Axiom 4 for $\forall$ and $\vdash \forall\varphi \rightarrow \mathcal{T}\varphi$
(2) $\forall\mathcal{T}\widehat{\mathcal{T}}\mathcal{T}\Diamond\Box\varphi \rightarrow \exists\mathcal{T}\widehat{\mathcal{T}}\mathcal{T}\Diamond\Box\varphi$	$\vdash \forall\varphi \rightarrow \exists\varphi$
(3) $\exists\mathcal{T}\widehat{\mathcal{T}}\mathcal{T}\Diamond\Box\varphi \rightarrow \exists\mathcal{T}\mathcal{T}\Diamond\Box\varphi$	The first step of the whole proof
(4) $\exists\mathcal{T}\mathcal{T}\Diamond\Box\varphi \rightarrow \exists\mathcal{T}\Diamond\Box\varphi$	Axiom T for $\mathcal{T}$
(5) $\forall\widehat{\mathcal{T}}\mathcal{T}\Diamond\Box\varphi \rightarrow \exists\mathcal{T}\Diamond\Box\varphi$	(1)(2)(3)(4)

The third step is to prove that

$$\vdash \exists\mathcal{T}\Diamond\Box\varphi \rightarrow \forall\widehat{\mathcal{T}}\mathcal{T}\varphi \ .$$

(1) $\exists\mathcal{T}\Diamond\Box\varphi \rightarrow \exists\mathcal{T}\varphi$	Theorem 3.5.2
(2) $\exists\mathcal{T}\varphi \rightarrow \exists\mathcal{T}\mathcal{T}\varphi$	Axiom 4 for $\mathcal{T}$
(3) $\exists\mathcal{T}\mathcal{T}\varphi \rightarrow \neg\exists\mathcal{T}\neg\mathcal{T}\varphi$	Axiom in ABS: $\vdash \mathcal{B}\varphi \rightarrow \neg\mathcal{B}\neg\varphi$
(4) $\neg\exists\mathcal{T}\neg\mathcal{T}\varphi \rightarrow \forall\widehat{\mathcal{T}}\mathcal{T}\varphi$	$\exists := \neg\forall\neg$, $\widehat{\mathcal{T}} := \neg\mathcal{T}\neg$
(5) $\exists\mathcal{T}\Diamond\Box\varphi \rightarrow \forall\widehat{\mathcal{T}}\mathcal{T}\varphi$	(1)(2)(3)(4)

At last we can conclude by applying what were achieved in the second and third step that

$$\vdash \forall\widehat{\mathcal{T}}\mathcal{T}\Diamond\Box\varphi \rightarrow \forall\widehat{\mathcal{T}}\mathcal{T}\varphi \ .$$

□

Therefore, together with the other direction, it follows that

$$\vdash_{ABS} \forall \widehat{\mathcal{T}}\mathcal{T} \Diamond \Box \varphi \leftrightarrow \forall \widehat{\mathcal{T}}\mathcal{T} \varphi \ .$$

By the soundness of ABS with respect to the class of AQA models (Theorem 2.5.7), we can conclude that

$$\vDash_{AQA} \forall \widehat{\mathcal{T}}\mathcal{T} \Diamond \Box \varphi \leftrightarrow \forall \widehat{\mathcal{T}}\mathcal{T} \varphi$$

which is equivalent to Theorem 3.4.6. In Appendix B, a model theoretical proof of Theorem 3.4.6 can be found. It is interesting to compare these two ways of proving the same theorem, which represent two different ways of understanding the same fact.

In the proof of $\vdash_{ABS} \forall \widehat{\mathcal{T}}\mathcal{T} \Diamond \Box \varphi \leftrightarrow \forall \widehat{\mathcal{T}}\mathcal{T} \varphi$, we have proved (the second step) that

$$\vdash_{ABS} \forall \widehat{\mathcal{T}}\mathcal{T} \Diamond \Box \varphi \to \exists \mathcal{T} \Diamond \Box \varphi \ .$$

Since $\forall \widehat{\mathcal{T}}\mathcal{T} \Diamond \Box \varphi$ expresses the definition of full-support belief in the class of AQA models according to Proposition 3.5.1 and $\exists \mathcal{T} \varphi$ expresses grounded belief, together with theorem 3.5.2, the formula tells us that full-support belief implies grounded belief. By noticing that

$$\vdash_{ABS} \forall \Diamond \Box \varphi \to \forall \widehat{\mathcal{T}}\mathcal{T} \Diamond \Box \varphi \ ,$$

we know that justified belief implies full-support belief.

So what full-support belief allows the agent to believe is more than justified belief but less than grounded belief (justified belief implies full-support belief and full-support belief implies grounded belief). The converse does not hold, as shown in the following example.

3.5.4. Example. Take the model in Example 2.3.14 and add a world 4 to the set of possible worlds. So $W = \{1,2,3,4\}$. Keep all the attack relations in the original model the same and add the following new attack relations:

- $P \leftarrowtail Q$ for all $P, Q \subseteq W$ such that $4 \in P$ and P includes less than 3 worlds and $P \cap Q = \varnothing$;
- $P \leftarrowtail Q$ for all $P, Q \subseteq W$ such that P is a singleton and Q includes exactly three worlds and $P \cap Q = \varnothing$
- $\varnothing$ is attacked by any other sets.

In this new model, the least fixed point is $\{P \subseteq W \mid \{1,2\} \subseteq P \text{ or } \{2,3\} \subseteq P \text{ or } \{1,3\} \subseteq P\}$. Reader can check that in this model the agent has grounded belief of $\{1,2\}$ but does not have full-support belief of $\{1,2\}$; the agent has full-support belief of $\{1,2,3\}$ but does not have justified belief of $\{1,2,3\}$.

Full-support belief satisfies all axioms and rules in the axiom system KD45. To be more precise, take a language $\mathcal{L}_f$ which includes only one modal operator $\mathcal{B}_f$ for full-support belief in the AQA models:

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathcal{B}_f\, \varphi \ .$$

$\mathcal{L}_f$ is a fragment of $\mathcal{L}_{\forall \mathcal{T} \Box}$. For all formulas $\varphi \in \mathcal{L}_f$,

$$\vDash_{AQA} \varphi \quad \text{if and only if} \quad \vdash_{KD45} \varphi \ .$$

The proof for the soundness and completeness is not hard, so we will not include all its details here. We only prove here the validity of conjunction closure for full-support belief in the class of AQA models.

3.5.5. Proposition.

$$\vDash_{AQA} \mathcal{B}_f\, \varphi \wedge \mathcal{B}_f\, \psi \to \mathcal{B}_f(\varphi \wedge \psi)$$

Proof:
The proof makes use of Proposition 3.5.1, which builds up the modal equivalence between $\mathcal{B}_f\, \varphi$ and $\forall \widehat{\mathcal{T}}\, \mathcal{T} \Diamond \Box\, \varphi$ in the class of AQA models. Then by $\vDash \forall \widehat{\mathcal{T}}\, \mathcal{T} \Diamond \Box\, \varphi \leftrightarrow \forall \widehat{\mathcal{T}}\, \mathcal{T}\, \varphi$, we only need to prove that

$$\vDash \forall \widehat{\mathcal{T}}\, \mathcal{T}\, \varphi \wedge \forall \widehat{\mathcal{T}}\, \mathcal{T}\, \psi \to \forall \widehat{\mathcal{T}}\, \mathcal{T}(\varphi \wedge \psi) \ .$$

Given an AQA model, assume that for all $f \in \mathsf{LFP}_\tau$ there is $f' \subseteq f$ such that $f' \in \mathsf{LFP}_\tau$ and $f' \subseteq [\![\varphi]\!]$, and there is $f'' \subseteq f$ such that $f'' \in \mathsf{LFP}_\tau$ and $f'' \subseteq [\![\psi]\!]$.

Take an arbitrary $t \in \mathsf{LFP}_\tau$. According to the assumption, there are $t' \subseteq t$ such that $t' \in \mathsf{LFP}_\tau$ and $t' \subseteq [\![\varphi]\!]$. By the assumption again, since $t' \in \mathsf{LFP}_\tau$, there is $t'' \subseteq t'$ such that $t'' \in \mathsf{LFP}_\tau$ and $t'' \subseteq [\![\psi]\!]$. Because $t'' \subseteq t' \subseteq [\![\varphi]\!]$ and $t'' \subseteq [\![\psi]\!]$, $t'' \subseteq [\![\varphi \wedge \psi]\!]$.

Therefore, for all $f \in \mathsf{LFP}_\tau$, we can find an argument $f' \subseteq f$ such that $f' \in \mathsf{LFP}_\tau$ and $f' \subseteq [\![\varphi \wedge \psi]\!]$. □

Therefore, full-support belief strikes a balance between justified belief and grounded belief, keeping the agent's beliefs fully consistent while allowing more to be believed.

3.5.6. Remark. Although in this section, we restrict our attention to the class of AQA models,

$$\vDash \forall \widehat{\mathcal{T}}\, \mathcal{T} \Diamond \Box\, \varphi \leftrightarrow \forall \widehat{\mathcal{T}}\, \mathcal{T}\, \varphi \tag{3.7}$$

actually also holds with respect to the whole class of topological argumentation models. But in the class of topological argumentation models, Theorem 3.4.6 cannot be expressed through 3.7. The difference is caused by a change concerning the semantic interpretation of $\Box$. In the class of AQA models, $\Box$ is not only an operator for *factive* evidence support but also a modal operator defined on the binary relation $\sqsubseteq_\tau$. They coincide in the class of AQA models. However, in the whole class of topological argumentation models, they do not coincide; $\Box$ does not function as the modal operator defined on the binary relation $\sqsubseteq_\tau$; and thus Proposition 3.5.1 does not hold with respect to the class of topological argumentation models.

3.6 Conclusion

In this Chapter, we take into consideration the role of default reasoning in shaping the agent's belief. Hence full-support belief strikes a balance between justified belief and grounded belief and relieves the tension between believing more and believing more consistently.

Although the family of notions of belief grows, the logical relations between them stays relatively simple. Moreover, they fit nicely into the strata of the agent's doxastic attitudes divided according to the level of consistency. We draw a map (Figure 3.3) to guide the tour among these notions of belief and the strata of the agent's doxastic attitudes. In the map, we use $\in$ to denote that a certain notion of belief belongs to a certain stratum and thus satisfies that level of consistency. And we use $\subseteq$ to denote the logical relation between different notions of belief. The arrows together with different types of reasoning indicate how the agent travels from one notion of belief to another one. Most importantly, the map is only valid with respect to the class of AQA models.

In the map, grounded belief and full-support belief are located between justified belief and evidence/arguments with respect to $\subseteq$. Especially, full-support belief is between justified belief and grounded belief with respect to $\subseteq$. While full-support belief keeps fully consistent as justified belief, it is proved to be a less stringent notion than justified belief. Although grounded belief is even less stringent than full-support belief, it fails to stay in the core of the doxastic earth.

The comparison of these different notions of belief is made possible by a unified framework – the logic of argument and belief. Note how justified belief, evidence/arguments, full-support belief and grounded belief can be expressed and reasoned about in the logic of argument and belief. Their logical relations are all encoded in the axiom system of ABL as demonstrated by the syntactic proofs in Section 3.5.

Semantically, in the logic of argument and belief, it is shown how these notions are connected with each other through different types of reasoning and different ways of applying these types of reasoning. It is reasoning that decides what the agent believes and how she believes it.

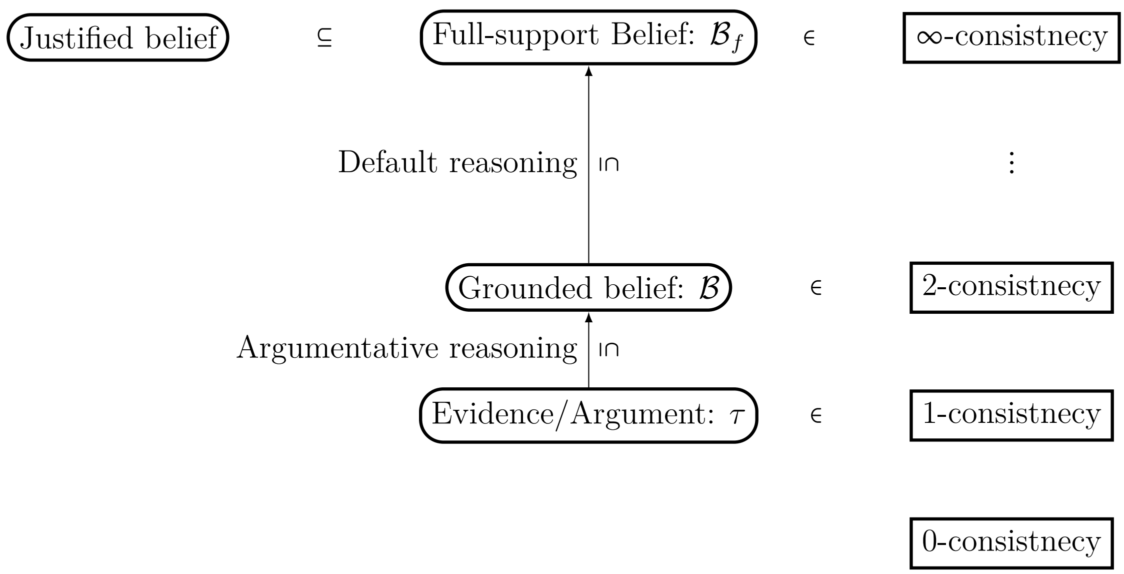

Figure 3.3: Map of different notions of belief with respect to the class of AQA models

Chapter 4

Group Belief through Argumentation

4.1 Introduction

In this chapter and the next chapter, we zoom out and study group belief. Distributed belief and everyone's belief are two notions of group belief in doxastic logic. The way distributed belief is defined implicitly assumes that the group can function as an astute logician by combining different group members' beliefs. However, different group members' beliefs can be incompatible with each other. Even though group members are willing to cooperate and learn from each other by incorporating others' beliefs, they still need to settle down the issue of incompatibility between their beliefs. The notion of distributed belief says nothing about how to deal with the incompatibility. Everyone's belief, on the other hand, avoids any inconsistency by requiring a consensus among group members – everyone agrees on the same opinion. However, the consensus is a very strong requirement, and the notion of everyone's belief says nothing about how such a consensus can be achieved.

Therefore, what we aim for in the following two chapters are notions of group belief satisfying the following two desiderata:

1. they should embody how conflicts between different group members' beliefs are resolved;

2. these notions of group belief should strike a balance between distributed belief and everyone's belief in the sense of being more consistent than distributed belief and allowing more to be believed than everyone's belief.

Debate or argumentation is one of the most natural ways of resolving conflicts between different group members' beliefs. The term "argumentation" in the previous two chapters refers to a way deliberation happens within someone's mind. The more literal meaning of "argumentation", nevertheless, refers to the activity of settling issues among a group of agents. In this chapter, we pursue a notion of group belief by analysing the structure of argumentation among the group members.

If we assume that a group can combine its group members' beliefs as efficiently and completely as a logically omniscient single agent combines her pieces of evidence, then the topological setting in the previous two chapters can be directly applied. The only thing we need to do is reinterpreting each evidence set in the basic evidence collection E_0 as a group member's set of doxastically accessible possible worlds. In this chapter, nevertheless, we do not follow the above assumption. We try to take into account the practical difficulties of coordinating different group members when logically combining group members' beliefs. Concerning the practical difficulties, the closure properties of the topology of arguments in the topological argumentation model are too idealistic to apply to a group's argumentation.[1]

The difficulties of coordination make the holistic view of reasoning and belief more problematic. Although for a single agent, by assuming that she is logically omniscient, it is acceptable to formalise her reasoning and belief globally rather than issue by issue[2], in the case of a group, such an idealisation seems unreasonable. If the group members have no idea what to argue for or against, it is impossible to have argumentation, not to mention the group belief based on argumentation.

The new setting in Section 4.2 – the multi-agent argumentation model – thus takes the differences between a group and a logically omniscient agent into account. One consequence of the differences is the inadequacy of a direct application of the semantics of grounded belief to group belief in the multi-agent argumentation model. We will explain the reason for the inadequacy and then propose a notion of group belief based on the greatest fixed point of the characteristic function in the multi-agent argumentation model, so-called "argumentation-based group belief". We will also show that the notion of argumentation-based group belief indeed satisfies the two desiderata.

A logic (the logic of group belief and argumentation, abbreviated to LGBA) is thus designed to express and reason about argumentation-based group belief and its relationship with each group member's belief in Section 4.3. We axiomatise LGBA and prove the soundness and completeness of the axiom system. At last, in Section Section 4.4 we discuss some further issues by comparing different notions of belief and conclude this chapter.

---

[1] This claim partially depends on how we merge models for separate agents into a multi-agent model. If we take product (Kurucz, 2007, Section 3) rather than fusion (Kurucz, 2007, Section 2), there might be a topological approach to the study of a group's argumentation, as indicated in the topological approach to the study of distributed knowledge and common knowledge in van Benthem and Sarenac (2004).

[2] For a single agent, her reasoning and belief are also issue-driven in practice. Some works take this factor into account, for example, van Benthem and Minică (2012).

Related works In the sense of following the approach of modal logic to the study of Dung's formal argumentation theory, the logic of group belief and argumentation is in a similar vein to a series of works by Davide Grossi (Grossi, 2010*b*,*a*, 2012, 2013; Grossi and van der Hoek, 2014). Technically speaking, LGBA has a flavour of a two-dimensional logic, which bears some similarity to the Facebook and epistemic logic developed in Seligman, Liu and Girard (2013). However, the differences are also significant:

1. LGBA is blended with ingredients from modal μ-calculus (Venema, 2008) – an operator for the greatest fixed point of the characteristic function;
2. the interaction between two dimensions is restricted in the syntax of LGBA, and we do not name each agent in the language.

Section 4.2 and 4.3 are based on Shi, Smets and Velázquez-Quesada (2017*b*, Section 4). In the same paper, readers can find more discussion about related works on belief and argumentation, although in Shi, Smets and Velázquez-Quesada (2017*b*) we focus on the individual agent's belief rather than group belief.

4.2 Multi-Agent Argumentation Frame

In this section, we will introduce the multi-agent argumentation model and propose a notion of group belief called "argumentation-based group belief".

4.2.1 Multi-agent argumentation frame (MAF)

Let us start with Dung's argumentation framework in Dung (1995, Definition 2).

4.2.1. Definition (Argumentation Framework). An argumentation framework is a pair

$$AF = (\mathsf{Ag}, \twoheadleftarrow)$$

where Ag is a set of nodes, and $\twoheadleftarrow \subseteq \mathsf{Ag} \times \mathsf{Ag}$ is a binary relation on Ag.

Dung interprets Ag in the argumentation framework as arguments and $\twoheadleftarrow$ the attack relation between arguments. In this section, instead of arguments, we take Ag in the argumentation framework as agents in a group.

To model each agent's belief on which group belief should hinge, we extend Dung's argumentation framework to multi-agent argumentation frame.

4.2.2. Definition (Multi-agent Argumentation Frame – Part One).
A *multi-agent argumentation frame* is a structure $\mathcal{F} = (W, \mathsf{Ag}, \{\twoheadleftarrow^P\}^{P\subseteq W}, f)$ where

- W is a non-empty set of possible worlds and Ag a non-empty set of agents;
- $\twoheadleftarrow^P \subseteq \mathsf{Ag} \times \mathsf{Ag}$ is an attack relation labeled by subsets of W $P \subseteq W$;

- $f : \mathsf{Ag} \to 2^W$ is a function assigning to each agent $s \in \mathsf{Ag}$ a subset of W such that $f(s) \neq \varnothing$ for all $s \in \mathsf{Ag}$.

The addition of a nonempty set of possible worlds W makes it possible to represent each agent's belief by assigning each agent a nonempty subset of W using function f. Moreover, the attack relation is indexed with a subset P of W, which is taken as an issue on whether P is the case. For different subset P of W, the attack relation $\twoheadleftarrow^P$ can be different. $s \twoheadleftarrow^P s'$ for $s, s' \in \mathsf{Ag}$ means agent s' attacks agent s on whether P is the case.

Compared to the topological argumentation model, each agent's belief in the multi-agent argumentation frame can be taken as a piece of evidence in the basic evidence collection E_0. There are two differences, reflecting the remarks we made in Section 4.1: in the multi-agent argumentation model

- there is no topology generated from the collection of each agent's belief; and
- the attack relation is issue-variant.

To make sense of the issue-variant attack relation among a group of agents, we need to impose the following three conditions on it:

4.2.3. Definition (Multi-agent Argumentation Frame – Part Two). The attack relation $\twoheadleftarrow^P$ in the multi-agent argumentation frame should satisfy the following three conditions:

1. $s \twoheadleftarrow^P s'$ if and only if $s \twoheadleftarrow^{\overline{P}} s'$;
2. if $s \twoheadleftarrow^P s'$, then
   (a) either $f(s) \subseteq P$ or $f(s) \subseteq \overline{P}$; and
   (b) $f(s) \subseteq P$ implies that $f(s') \subseteq \overline{P}$;
3. if $s \twoheadleftarrow^P s'$ and $f(s) \subseteq Q \subseteq P$, then $s \twoheadleftarrow^Q s'$.

The first condition captures the intuition that a debate on whether P is the case is also a debate on whether $\overline{P}$ is the case. Condition 2(a) requires that if an agent has no stance on whether P is the case, then the agent would not be attacked; Condition 2(b) says (together with 1 and 2(a)) that, when an agent attacks another agent, they should hold the opposite stances on the issue at hand. Note that it does not hold in reverse, given these conditions, which means that there can be two agents who hold the opposite stances but they do not attack each other. So it is quite different from the requirements on the attack relation in the topological argumentation model. The last condition states that if s' attacks s on her belief of P, then s' should also attack s on her belief of Q which is logically stronger than P.

Given a multi-agent argumentation frame , let

$$\mathsf{Def}^P(s) = \{s' \in \mathsf{Ag} \mid \text{there is a path } s \twoheadleftarrow^P s_1 \twoheadleftarrow^P \ldots \twoheadleftarrow^P s_n \twoheadleftarrow^P s' \text{ with even steps}\}$$

and

$$\mathsf{Att}^P(s) = \{s' \in \mathsf{Ag} \mid \text{there is a path } s \twoheadleftarrow^P s_1 \twoheadleftarrow^P \ldots \twoheadleftarrow^P s_n \twoheadleftarrow^P s' \text{ with odd steps}\}$$

for each $s \in \mathsf{Ag}$.

4.2.4. Definition (Uncontroversial Attack Relation). $\twoheadleftarrow^P$ in a multi-agent argumentation frame is uncontroversial if and only if

$$\mathsf{Def}^P(s) \cap \mathsf{Att}^P(s) = \varnothing$$

for all $s \in \mathsf{Ag}$.

An important implication of Condition 1 and 2 is that

4.2.5. Proposition. *For all $P \subseteq W$, $\twoheadleftarrow^P$ in the multi-agent argumentation frame is uncontroversial.*

The following example illustrates how the multi-agent argumentation frame can be used to model some scenarios.

4.2.6. Example. A group of three students cooperates on a home assignment from a zoology course, which asks them to classify an animal in a picture. Each student in the group comes up with his answer.

- s_1: The animal is a bird.
- s_2: The animal is a mammal.
- s_3: The animal is a reptile.

After a round of argumentation, given that s_1 is attacked by both s_2 and s_3, while s_2 and s_3 attack each other, we can represent the scenario by the following multi-agent argumentation frame

$$(\{b, m, r\}, \{s_1, s_2, s_3\}, \{\twoheadleftarrow^P\}^{P \subseteq W}, f)$$

with the attack relation given by

$$\begin{aligned} &\twoheadleftarrow^{\{b\}} = \twoheadleftarrow^{\{m,r\}} := \{(s_1, s_2), (s_1, s_3)\}, && \twoheadleftarrow^{\{m\}} = \twoheadleftarrow^{\{b,r\}} := \{(s_1, s_2), (s_3, s_2), (s_2, s_3)\} \\ &\twoheadleftarrow^{\varnothing} = \twoheadleftarrow^{\{b,m,r\}} := \varnothing && \twoheadleftarrow^{\{r\}} = \twoheadleftarrow^{\{b,m\}} := \{(s_1, s_3), (s_3, s_2), (s_2, s_3)\} \end{aligned}$$

and

$$f(s_1) = \{b\}, \qquad f(s_2) = \{m\}, \qquad f(s_3) = \{r\}.$$

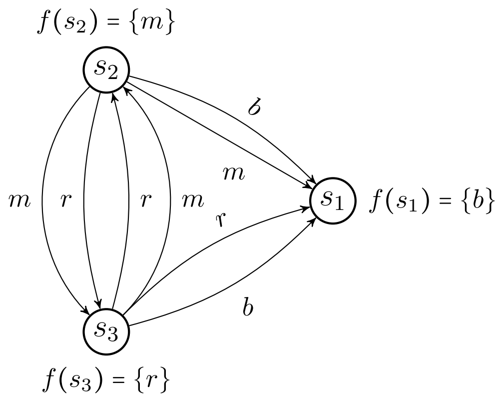

Figure 4.1: The multi-agent argumentation frame for Example 4.2.6.

Figure 4.1 illustrates the multi-agent argumentation model (for simplicity, not all attack relations are drawn).

Note that we intentionally leave out the details of how the attack relations between agents are pinned down in the example. The question is out of our concern in this thesis and thus is not something can be modelled by the multi-agent argumentation frame. We just assume that the attack relations are given and satisfy the three conditions.

4.2.2 Argumentation-based group belief

Given an issue of whether P is the case and the attack relation with respect to this issue among a group of agents, which stance, P or $\overline{P}$, should the group follow?

In Chapter 2, we define a single agent's grounded belief based on the least fixed point semantics. Does the grounded semantics also work for argumentation of a group of agents?

4.2.7. Definition (Characteristic function and least/greatest fixed point). Given a multi-agent argumentation frame, let $d^P : 2^{\mathsf{Ag}} \to 2^{\mathsf{Ag}}$ denote the characteristic function with respect to $\twoheadleftarrow^P$:

$$d^P(A) = \{s \in \mathsf{Ag} \mid s \text{ is defended by } A \subseteq \mathsf{Ag} \text{ with respect to } \twoheadleftarrow^P\} \ ;$$

and let $\mathsf{LFP}^P/\mathsf{GFP}^P$ denote the least/greatest fixed point of function d^P.

A direct application of the semantics of grounded belief leads us to the following definition of group belief: the group believes P in $\mathcal{F}$ if and only if there is an agent $s \in \mathsf{LFP}^P$ such that $f(s) \subseteq P$. Under this definition of group belief, however, it is possible for the group to believe both P and $\overline{P}$ which is undesirable in the sense of being against the 2-consistency. The reason is the possibility of a lack

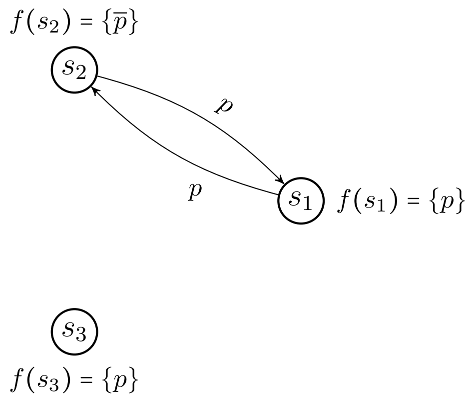

Figure 4.2: The multi-agent argumentation frame for Example 4.2.8.

of attack relation between two agents who have the opposite stances, which thus allows two agents with the opposite beliefs to belong to LFP^P at the same time. To avoid the problem, how about modifying the definition as follows: the group believes P if and only if there is an agent s belonging to LFP^P such that $f(s) \subseteq P$ and there is no agent s' belonging LFP^P such that $f(s') \subseteq \overline{P}$?

The following example shows that the modified definition is still unsatisfactory in the sense of underestimating agents outside the least fixed point.

4.2.8. Example. Consider the multi-agent argumentation frame

$$(W = \{p, \overline{p}\}, \mathsf{Ag} = \{s_1, s_2, s_3\}, \{\twoheadleftarrow^P\}^{P \subseteq W}, f)$$

where

$$\twoheadleftarrow^{\{p\}} = \twoheadleftarrow^{\{\overline{p}\}} = \{(s_1, s_2), (s_2, s_1)\}$$

and $f(s_1) = f(s_3) = \{p\}$ and $f(s_2) = \{\overline{p}\}$. The frame is illustrated in Figure 4.2.

Indeed, there is an agent s_3 in $\mathsf{LFP}^{\{p\}}$ who believes $\{p\}$ and there is no agent in $\mathsf{LFP}^{\{p\}}$ who believes $\{\overline{p}\}$. However, agent s_2, who is not in the least fixed point, believes $\{\overline{p}\}$ and she can defend herself from any other attacker. The membership of s_3 in the least fixed point is partly resulted from the absence of attack relation between s_3 and s_2.[3]

Example 4.2.8 shows that the membership of the least fixed point in the multi-agent argumentation frame is not decisive for group belief. What matter are the following concepts.

4.2.9. Definition. Given a multi-agent argumentation frame,

- a subgroup of agents $A \subseteq \mathsf{Ag}$ is P-admissible if and only if A is conflict-free (there is no $s, s' \in A$ such that $s \twoheadleftarrow^P s'$) and can defend itself ($A \subseteq d^P(A)$) with respect to $\twoheadleftarrow^P$;

[3] Note that this is impossible in the topological argumentation model because when the intersection of two pieces of evidence is empty, there must be an attack relation between them.

- an agent $s \in \mathsf{Ag}$ is acceptable if and only if s belongs to a P-admissible subgroup of agents for some $P \subseteq W$.

Our definition of group belief in this Chapter is thus based on these two concepts.

4.2.10. Definition (Argumentation-based group belief). Given a multi-agent argumentation frame $\mathcal{F}$, the group has argumentation-based belief of P in $\mathcal{F}$ if and only if there is an acceptable agent believing P and there is no acceptable agent believing $\overline{P}$.

The definition has nothing to do with the least fixed point directly but rather focuses on the existence of acceptable agents believing a certain proposition and the non-existence of acceptable agents believing the negation of the proposition. It thus weighs two sides of a binary issue equally.

However, argumentation-based group belief is closely related to the greatest fixed point.

4.2.11. Proposition. *Given a multi-agent argumentation frame $\mathcal{F}$, the group has argumentation-based belief of P in $\mathcal{F}$ if and only if there is an agent $s \in \mathsf{GFP}^P$ such that $f(s) \subseteq P$ and there in no agent $s' \in \mathsf{GFP}^P$ such that $f(s') \subseteq \overline{P}$.*

The proof of this proposition follows immediately from the next lemma.

4.2.12. Lemma. *Given a multi-agent argumentation frame $\mathcal{F}$, there is an P-admissible subgroup of agents $A \subseteq \mathsf{Ag}$ such that $s \in A$ if and only if $s \in \mathsf{GFP}^P$, for all agent $s \in \mathsf{Ag}$ and all $P \subseteq W$.*

Proof:
See Appendix C.1. □

Proposition 4.2.11 facilitates a way of characterising argumentation-based group belief in the logic of group belief and argumentation studied in the next section. Before introducing the logic, we bring up the following two facts about the closure properties of argumentation-based group belief.

4.2.13. Proposition. *Given a MAF $\mathcal{F}$ and any $P, Q \subseteq W$,*

1. *if group has argumentation-based belief of P and $P \subseteq Q$, then the group also has argumentation-based belief of Q;*
2. *even if the group has argumentation-based belief of P and has argumentation-based belief of Q, it may not have argumentation-based belief of $P \cap Q$.*

Proof:
The proof is essentially the same as the proof of Shi, Smets and Velázquez-Quesada (2017*b*, Corollary 3.1) which follows from Shi, Smets and Velázquez-Quesada (2017*b*, Proposition 3.2). So we do not repeat it here. Intuitively, the

failure of closure under intersection for argumentation-based group belief is not so surprising, considering that for some MAFs, $\{f(i) \mid i \in \mathsf{Ag}\}$ is not closed under conjunction. $\square$

4.3 Logic of Group Belief and Argumentation

In this section, we present a sound and complete logic for characterising a group member's belief, attack relations between different agents and argumentation-based group belief in the multi-agent argumentation model.

4.3.1 Syntax

4.3.1. Definition. Let At be a non-empty set of atomic propositions. $\mathcal{L}_{\alpha\beta}$ is the language generated by the following grammar:

$$\alpha ::= \top \mid p \mid \neg\alpha \mid \alpha \wedge \alpha \mid \boxminus\alpha \mid \boxdot\beta$$
$$\beta ::= \top \mid \Box\alpha \mid \neg\beta \mid \beta \wedge \beta \mid [\alpha]\beta \mid \mathsf{Gfp}^{\alpha}$$

where $p \in \mathsf{At}$. Symbols $\diamondminus, \Diamond, \langle\alpha\rangle$ and $\bot$ are the abbreviations of $\neg\boxdot\neg, \neg\Box\neg, \neg[\alpha]\neg$ and $\neg\top$, respectively.

The language comprises two parts which are partially intertwined with each other.

The α part of the language, which we will call "α-formulas", is mainly defined on the dimension of possible worlds, in order to describe facts about possible worlds. For example, $\boxminus$ is a universal operator quantifying over possible worlds; analogously, $\boxdot$ is used to talk about the whole group of agents. Note that $\boxdot\beta$ is an α-formula. For example, the sentence "all the group members believe that they are logicians" is a fact holding in a certain possible world.

The β part, which we will call "β-formulas", is mainly defined on the dimension of agents to express facts about a referred agent in argumentation. For example, formulas of the form $\Box\alpha$ state that the referred agent believes α. Formulas of the form $[\alpha]\beta$ state that all agents who directly attack the referred agent on the issue of whether α is the case satisfy β. Notice the positions of α and β in formulas of the form $[\alpha]\beta$. Gfp^{α} expresses that the referred agent is acceptable in the argumentation about α.

There are several reasons for the restriction on the interactions between α-formulas and β-formulas.

1. Some of the interactions between the two dimensions do not make much sense. For example, in $\boxdot\boxdot\beta$, $\boxdot\beta$ expresses a fact about possible worlds (it is an α-formula), so we cannot use it to describe agents.

2. We do not take everything to be an issue of argumentation. For example, facts about argumentation itself which are described by β-formulas.

3. Technically such a restriction makes a complete axiomatization feasible.

4.3.2 Semantics

By adding a valuation function $V : \mathsf{At} \to 2^W$ to the multi-agent argumentation frame $\mathcal{F}$, we get the multi-agent argumentation model $\mathcal{M} = (\mathcal{F}, V)$, where formulas in $\mathcal{L}_{\alpha\beta}$ can be evaluated.

Let

$$[\![\alpha]\!]_{\mathcal{M}} \coloneqq \{w \in W \mid \mathcal{M}, (w, s) \vDash \alpha \text{ for all arguments } s \in \mathsf{Ag}\}$$

(the subscript $\mathcal{M}$ will be omitted whenever possible). The truth of $\varphi \in \mathcal{L}_{\alpha\beta}$ is defined as follows:

4.3.2. Definition. Given a multi-agent argumentation model $\mathcal{M} = (W, \mathsf{Ag}, \{\leftarrowtail^P\}^{P \subseteq W}, f, V)$,

- $\mathcal{M}, (w, s) \vDash \top$
- $\mathcal{M}, (w, s) \vDash p$ **iff** $w \in V(p)$
- $\mathcal{M}, (w, s) \vDash \neg\varphi$ **iff** $\mathcal{M}, (w, s) \nvDash \varphi$
- $\mathcal{M}, (w, s) \vDash \varphi \wedge \varphi'$ **iff** $\mathcal{M}, (w, s) \vDash \varphi$ and $\mathcal{M}, (w, s) \vDash \varphi'$
- $\mathcal{M}, (w, s) \vDash \boxminus\alpha$ **iff** for all $w' \in W$, $\mathcal{M}, (w', s) \vDash \alpha$
- $\mathcal{M}, (w, s) \vDash \boxbar\beta$ **iff** for all $s' \in \mathsf{Ag}, \mathcal{M},\ (w, s') \vDash \beta$
- $\mathcal{M}, (w, s) \vDash \Box\alpha$ **iff** $f(s) \subseteq [\![\alpha]\!]$
- $\mathcal{M}, (w, s) \vDash [\alpha]\beta$ **iff** for all $s' \in \mathsf{Ag}$ such that $s \leftarrowtail^{[\![\alpha]\!]} s'$, $\mathcal{M}, (w, s') \vDash \beta$.
- $\mathcal{M}, (w, s) \vDash \mathsf{Gfp}^\alpha$ **iff** $s \in \mathsf{GFP}^{[\![\alpha]\!]}$.

We say a formula φ is satisfied in a multi-agent argumentation model $\mathcal{M}$ if there is a pair (w, s) in $\mathcal{M}$ such that $\mathcal{M}, (w, s) \vDash \varphi$. A formula φ is valid in $\mathcal{M}$ ($\mathcal{M} \vDash \varphi$) if for all pairs (w, s) in $\mathcal{M}$ we have $\mathcal{M}, (w, s) \vDash \varphi$. And φ is valid in the whole class of multi-agent argumentation models ($\vDash \varphi$) if it is valid in every multi-agent argumentation model.

Proposition 4.2.11 facilitates the following way of expressing argumentation-based group belief in the logic of group belief and argumentation:

$$\mathcal{B}_c\, \alpha \coloneqq \diamondminus(\Box\alpha \wedge \mathsf{Gfp}^\alpha) \wedge \neg \diamondminus (\Box\neg\alpha \wedge \mathsf{Gfp}^{\neg\alpha})$$

Except for the logical properties mentioned at the end of Section 4.2, $\mathcal{B}_c$ also satisfies the following properties.

4.3.3. FACT.

$$\models \mathcal{B}_c\,\alpha \to \neg\,\mathcal{B}_c\,\neg\alpha, \qquad \models \alpha \text{ implies } \models \mathcal{B}_c\,\alpha$$
$$\models \mathcal{B}_c\,\alpha \to \mathcal{B}_c\,\mathcal{B}_c\,\alpha, \qquad \models \neg\,\mathcal{B}_c\,\alpha \to \mathcal{B}_c\,\neg\,\mathcal{B}_c\,\alpha.$$

4.3.3 Axiom system

The following list of formulas of rules constitutes the axiom system (called GBAS) for the logic of group belief and argumentation. Axioms for Gfp^α indicate that this operator amounts to the greatest fixed point of $d^{[\![\alpha]\!]}$.

- All the propositional tautologies.
- Modus ponens
- S5 and Necessitation rule for $\boxminus$
- For $\boxbar$:

  K $\vdash \boxbar(\beta \to \beta') \to (\boxbar\beta \to \boxbar\beta')$

  D $\vdash \neg\,\boxbar\,\bot$

  N If $\vdash \beta$, then $\vdash \boxbar\beta$

- For $\Box$:

  K $\vdash \Box(\alpha \to \alpha') \to (\Box\alpha \to \Box\alpha')$

  D $\vdash \neg\,\Box\,\bot$

  N If $\vdash \alpha$, then $\vdash \Box\alpha$

- For $[\alpha]$:

  K $\vdash [\alpha](\beta \to \beta') \to ([\alpha]\beta \to [\alpha]\beta')$

  N $\vdash \beta$ implies $\vdash [\alpha]\beta$

- For Gfp^α:

  Unfold $\vdash \mathsf{Gfp}^\alpha \to [\alpha]\langle\alpha\rangle\,\mathsf{Gfp}^\alpha$

  R $\vdash \beta \to [\alpha]\langle\alpha\rangle\beta$, then $\vdash \beta \to \mathsf{Gfp}^\alpha$

- Interaction between $\boxbar$ and $\boxminus$

  $\boxplus$**1** $\vdash \boxbar\beta \to \boxminus\,\boxbar\,\beta$

  $\boxplus$**2** $\vdash \neg\,\boxbar\,\beta \to \boxminus\neg\,\boxbar\,\beta$

- Interaction between $\boxminus$, $\boxbar$, $\Box$ and $[\alpha]$

I1 $\vdash \boxminus\alpha \leftrightarrow \boxbar \Box \boxminus\alpha$

I2 $\vdash \boxbar\beta \rightarrow \boxbar[\alpha]\beta$

- Interaction between $\Box$ and $[\alpha]$

1 $\vdash [\alpha]\beta \leftrightarrow [\neg\alpha]\beta$

2a $\vdash \langle\alpha\rangle\top \rightarrow \Box\alpha \vee \Box\neg\alpha$

2b $\vdash \Box\alpha \rightarrow [\alpha] \Box \neg\alpha$

3 $\vdash \Box \boxminus (\alpha \rightarrow \alpha') \wedge \Box\alpha \wedge \langle\alpha'\rangle\beta \rightarrow \langle\alpha\rangle\beta$

Since $\boxbar\beta \rightarrow \beta$ and $\boxbar \boxbar \beta$ are not expressible in our language, axioms **T**, **4** and **5** do not hold for $\boxbar$, which explains why we need axioms $\boxplus$**1**, $\boxplus$**2**, **I1** and **I2** to characterize the relationship between $\boxbar$ as a universal operator and other operators.

Axioms **1**, **2a**, **2b** and **3** correspond to frame conditions 1, 2(a), 2(b) and 3 in Definition 4.2.3, respectively.

The axioms for Gfp^{α} are special cases of the general greatest fixed point operator (Kozen, 1983; Venema, 2008): the **unfold** axiom says that Gfp^{α} is a postfix point of $[\alpha]\langle\alpha\rangle$, and rule **R** says that Gfp^{α} is the greatest postfix point

4.3.4. Theorem. *The system GBAS is sound and weakly complete with respect to the class of multi-agent argumentation models*

The details of the completeness proof can be found in Appendix C.2. Here we only outline the proof and highlight some points.

The completeness proof follows the standard procedure of constructing a model by making use of maximally consistent sets of formulas and proving the truth lemma which builds the equivalence between the satisfiability of some given formulas in the model on a maximally consistent set and their membership in the set.

During the proof, the first twist we meet is about constructing the pairs of possible worlds and agents in the desired model. Notice that although the multi-agent argumentation model is two-dimensional, the syntax restricts the interaction between these two dimensions. For example, strings as $\boxbar\beta \rightarrow \beta$ are not formulas in our language: $\boxbar\beta$ is an α-formula, but β is not. So $\{\boxbar\beta, \neg\beta\}$ is a consistent set in our logic, even though no pair (w, s) in a multi-agent argumentation model satisfies both $\boxbar\beta$ and $\neg\beta$. This is a double-edged sword. On the one hand, it gives us the flexibility to construct maximal consistent sets for α-formulas and β-formulas separately and put them together in the model. On the other hand, we have to use some devious ways to ensure the satisfiability of the maximal consistent sets.

The second twist is about constructing the attack relations labelled by all subsets of possible worlds in the model. We may get no information about the attack relation with a certain label directly from the formulas which are generated from the subformulas of a given formula ϕ. This means we can only construct the model partially based on the given syntactical information. Therefore, we have to prove that this kind of partial model in the proof can be extended to a full and real multi-agent argumentation model. Moreover, the extended model should be modally equivalent to the original model with respect to all the given formulas.

At last, the proof for the case of Gfp^{α} in the truth lemma is worth some attention. It is not as straightforward as other cases. The readers can find the details of how axiom **Unfold**, and the inference rule **R** are applied in the proof.

4.4 Discussion and Conclusion

Argumentation-based group belief meets both requirements specified at the beginning of this chapter. It explicitly prescribes a way of treating the inconsistency between different group members' beliefs. Compared with distributed belief, it is 2-consistent and thus more consistent. Compared with everyone's belief, it allows more to be believed in the sense of being strictly weaker than everyone's belief in LGBA (the latter implies the former).

Despite these positive sides compared with distributed belief and everyone's belief, argumentation-based group belief is still not completely satisfactory concerning its level of consistency. Is there any fix for this problem for argumentation-based group belief as full-support belief for grounded belief?

We do not have a concrete proposal in reply to the above question. Nonetheless, a brief comparative analysis of argumentation-based belief and grounded belief may help point to a direction worth further exploration. First, as we stated at the beginning of Section 4.1, argumentation-based group belief does not assume that the group would first pool each member's belief together and extract as much consistent information as possible. Neither does it require the group to do this after the conflict is resolved for each proposition. On the contrary, grounded belief, as a notion of belief for the single agent, combines logical reasoning and argumentational reasoning rather than separating them. The argumentational reasoning is done on the basis of the topology of evidence/arguments, which is achieved by the agent's logical reasoning. Second, by labeling the attack relation with different binary issues, argumentation-based group belief is defined issue by issue rather than holistically as grounded belief is based on argumentational reasoning. These two points – lack of a proper way of combining logical reasoning and argumenational reasoning and being issue-driven – seem to cause more difficulties for achieving full consistency. A better understanding of the two points might be helpful for finding a solution to the problem of lacking full consistency. Moreover, a further study of this problem may lead us to a richer taxonomy of notions of group belief as what we achieved in Chapter 3 for single-agent belief based on different types of reasoning and their combination.

Besides the logical study of argumentation-based group belief, the logic of group belief and argumentation is another contribution of this chapter. Different from the logic of argument and belief in Section 2.5, LGBA makes the dimension of argumentation explicit rather than hide it from an operator for belief. Adding an operator for the greatest fixed point in the language of the logic and labeling the attack relation with different issues add up to the complication of the logic. By restricting the interactions between the two dimensions of the logic, we manage to achieve a sound and complete axiom system. For our purpose in this chapter, the restriction is reasonable. However, from a technical point of view, a real two-dimensional logic based on our setting without imposing any restriction is worth a further study.

Chapter 5

Potential Group Belief

> P'ang Ts'ung was being sent with the heir-apparent as a hostage to Han-tan.
>
> He spoke to the King of Wei and said:'If now one man said that there was a tiger in the market-place, would Your Majesty believe it?'
>
> The King said 'No.'
>
> 'If two men said that there was a tiger in the market-place, would Your Majesty believe it?'
>
> The King said:'I should be suspicious.'
>
> 'If three men said that there was a tiger in the market-place, would Your Majesty believe it?'
>
> The King said:'I should believe it.'
>
> P'ang Ts'ung said:'It is clear that there is no tiger in the market-place, and yet three men's words would make a tiger...'
>
> (*Records of Warring States*)[1]

5.1 Introduction

In the previous Chapter, we force the group members into an arena and initiate the gladiator combat. It is a tough, hard and immediate way of resolving the conflicts between group members and deciding the group's doxastic attitude toward some issues. However, this notion of group belief seems a little ad hoc in the sense of relying on the explicit specification of the attack relation between agents, which is not inherent to the group and its members. The group would not have any beliefs unless a debate is organized. Hence the resulting notion of group belief would not be able to trace and reflect the change of the individual group members' beliefs very well. In a nutshell, the argument-based group belief is not as intrinsic to the group as the single-agent belief to the single agent.

[1]The translation is excerpted from https://lib.hku.hk/bonsall/zhanguoce/index1.html

In this Chapter, we set the gladiators free so that they interact with each other in a gentle, soft and evolving way. They are allowed to make friends and choose whom to trust and to follow. Mostly they have different opinions. Meanwhile, they are exposed to each other's influence through their interaction. The more people surrounding you have the same opinion, the higher is the chance that you would follow them, just as the King in the story from the ancient Chinese text *Records of Warring States*.

So instead of picking out some winners' belief as the group's belief by an external mechanism (argumentation, voting and so on), we observe and explore the way in which the group members influence each other and how the group members' beliefs would evolve according to the influence flow in their current spontaneous interaction. The trend of the evolution hints at the group's potential doxastic state.

Rather than ambitiously taking all the components involved in the social interaction into account, we start with a simplified characterisation, abstracting away some details (communication, persuasion and so on). The only thing about the interaction we keep and specify in the model is how much one member trusts another. According to the current beliefs of the ones she trusts, the agent decides how to change her belief. So the whole group is undergoing a process of belief change. We will introduce in Section 5.2 two existing models in the literature which characterise the process. In both models, this process is treated as a deterministic process. Our work differs from these two models by taking the process as an indeterministic process, which is a Markov process in essence, as we will show in Section 5.3.

In Section 5.5, we will introduce the notion of potential group belief based on the trend of the process, which is explored in detail in Section 5.4. Our analysis will make use of results from Markov chain theory. Furthermore, we study the logic of potential group belief in Section 5.6 and Section 5.7. Technically, the logic is closely related to those logics of qualitative probability studied in Segerberg (1971), Gärdenfors (1975), Holliday and Icard III (2013) and van Eijck and Li (2017). For potential group belief, the conjunction rule fails again. We propose a way of saving the failure in which a new notion of group belief emerges in Section 5.8. In the same section, we also discuss its connection with the theory of judgement aggregation. In Section 5.9, we conclude the whole chapter, emphasising some important features of potential group belief.

This Chapter is multi-flavoured, as the overview of its structure indicates. The notion of potential group belief is likely to the taste of some epistemologists and epistemic logicians, who are interested in group notions of belief and knowledge, for example, common belief/knowledge and distributed belief/knowledge. The key idea behind this new notion of belief – group belief as group tendency to a certain stable state – seems quite relevant to the discussion about group agency (List and Pettit, 2011; Bratman, 2014). For some mathematicians, the alternative proof of the convergence result for the regular Markov chain in Appendix D.2 and the relation between the convergence results about the absorbing Markov chain and the regular Markov chain revealed in our analysis about potential group belief and crystallised in Theorem 5.5.9 may be worth some attention. Others, for example, social choice theorists, are also expected to taste some familiar flavours. All the results presented as propositions, lemmas, theorems and corollaries, unless their sources are explicitly stated, are newly proved.

5.2 DeGroot Model and Threshold Model

In this section, we introduce the DeGroot model (DeGroot, 1974) and the threshold model [2] which are used to model the group members' belief change under social influence.

Both of these models include a trust matrix specifying how much one agent trusts another. Let G be a *finite* group of agents. Each agent distributes her trust among the agents in the group, including herself. The more agent i trusts agent j about a certain issue, the more influence agent j has on agent i about the issue. How much trust agent i puts on agent j is represented by a real number $\mathbf{T}_{ij}$. In both models, it is required[3] that for all $i, j \in G$,

$$\mathbf{T}_{ij} \in [0, 1]$$

and

$$\sum_j \mathbf{T}_{ij} = 1\,.$$

We can organize these real numbers into the form of a matrix, called a "trust matrix" and denoted by $\mathbf{T}$.

5.2.1. Example. Given a group of agents $\{1, 2, 3\}$, it can have the following trust matrix

$$\mathbf{T} = \begin{bmatrix} 0 & 1 & 0 \\ 0.4 & 0 & 0.6 \\ 1 & 0 & 0 \end{bmatrix}$$

where $\mathbf{T}_{12} = 1$ and $\mathbf{T}_{23} = 0.6$, indicating that agent 1 only trusts agent 2 and agent 2 puts weight 0.6 on agent 3.

The essential difference between the DeGroot model and the threshold model lies in their representations of each agent's belief and how each agent updates her belief according to the group members' beliefs.

---

[2] The threshold model is studied in a wide range of areas, for example, social networks (Poljak and Sůra, 1983), computer science (Goles Ch, 1985), logic (Liu, Seligman and Girard, 2014; Baltag, Christoff, Rendsvig and Smets, 2016) and social choice (Grandi, Lorini and Perrussel, 2015).

[3] This requirement is not necessary for the threshold model, we state it here for both models just for a unifying treatment.

In the DeGroot model, each agent's belief is represented by her subjective probability of a given proposition being true. If there are n agents in the group G, then the agents' subjective probabilities in this group can be represented by an n-dimensional vector with its ith entry representing the ith agent's subjective probability.

5.2.2. Example. Given a group of three agents, if the first agent assigns probability 0.2 to a given proposition, the second agent assigns 0.3 and the third one assigns 0.9, then their beliefs can be represented by the following vector.

$$\mathbf{c} = \begin{bmatrix} 0.2 \\ 0.3 \\ 0.9 \end{bmatrix}$$

In the threshold model, only the binary qualitative belief is dealt with, that is, the agent either believes or does not believe the given proposition, there is no in-between state. For a group of n members, its members' beliefs can also be written as an n-dimensional vector.

5.2.3. Example. Given a group of three agents, if the first agent believes the given proposition, the second one does not, and the third one believes it, then it can be represented by the following vector.

$$\mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

where 1 means the agent believes the proposition while 0 means the agent does not believe the proposition (but not necessarily believes its negation).

For clarity, we will distinguish between two terms "credence vector" and "binary belief vector". The former is a vector with real numbers between 0 and 1 as items; the latter is a vector with 0 or 1 as items. The DeGroot model deals with the credence vector, and the threshold model only deals with the binary belief vector.

The difference in their representations of each agent's belief results in different updating methods adopted by the two models. In the DeGroot model, given the group members' credence vector $\mathbf{c}^{(0)}$ and the trust matrix $\mathbf{T}$, the group members' credence vector after one round of updating can be computed by

$$\mathbf{c}^{(1)} = \mathbf{T}\mathbf{c}^{(0)}\,.$$

Note that we use the superscript $c^{(n)}$ to indicate that it is the credence vector for the group after n rounds of updating.

Given an agent i in the group G, her credence to the given proposition after one round of update $\mathbf{c}_i^{(1)}$ (the ith entry of the credence vector) is the weighted sum of all group members' initial credence $\sum_j \mathbf{T}_{ij}\mathbf{c}_j^{(0)}$. To be more general, for all integers $n > 0$,

$$\mathbf{c}^{(n)} = \mathbf{T}\mathbf{c}^{(n-1)} .$$

Or equivalently,

$$\mathbf{c}^{(n)} = \mathbf{T}^n\mathbf{c}^{(0)}$$

where $\mathbf{T}^n$ is the nth power of the trust matrix $\mathbf{T}$.

The same updating method cannot be used in the threshold model since $\mathbf{Tb}$ is not necessarily a binary belief vector when $\mathbf{b}$ is. So in the threshold model, the update rule is the threshold rule as defined subsequently. Let $\mathbf{r}$ be a vector each of whose items is a real number, denoting the threshold for the corresponding group member. Let $F : \mathbb{R}^m \mapsto \{0,1\}^m$ be a function mapping strictly positive items of vectors in $\mathbb{R}^m$ to 1 in a binary vector and mapping the other items to 0 in a binary vector. Then

$$\mathbf{b}^{(n+1)} = F(\mathbf{T}\mathbf{b}^{(n)} - \mathbf{r})$$

This rule says that if the weighted sum $\sum_{j\in G}\mathbf{T}_{ij}\mathbf{b}_j^{(n)}$ is larger than agent i's threshold $\mathbf{r}_i$, then agent i would turn to believe the given propositions in the next step; otherwise, the agent i would not believe it in the next step. In this way, the agents in the group keep their beliefs qualitative.

Compared to the update rule in the DeGroot model, the threshold rule takes one more step; it compares the weighted sum with the threshold and decides whether the agent should change her belief or not. Note that the threshold r_i can be different with respect to different agents.

Given these different update rules, some interesting questions about the dynamics of the two models are studied in the literature (DeGroot, 1974; Poljak and Sůra, 1983; Goles Ch, 1985; Golub and Jackson, 2010). For example, the following theorem can be derived by applying Theorem 3.2 in Goles Ch (1985) to a given threshold model:

5.2.4. THEOREM. *Given a threshold model, if the trust matrix* $\mathbf{T}$ *is symmetric, then every sequence* $\mathbf{b}^{(0)}, \mathbf{b}^{(1)}, \ldots$ *generated by the iteration of applying the threshold rule to* $\mathbf{T}$ *would either converge to a fixed point or oscillate between two states after a certain number of steps in the sequence.*

We will see more theorems of a similar kind about the dynamics of the DeGroot model when we discuss the dynamics of an indeterministic binary DeGroot model in the next section. We will also show that there is a close connection between the new indeterministic model and the DeGroot model.

5.3 Indeterministic Binary DeGroot Model

In Section 5.2, it is shown that the opinion pool, namely the linear combination of the form $\mathbf{Tc}$ or $\mathbf{Tb}$, is the common step taken in the updating mechanisms of the

DeGroot model and the threshold model. Despite some restrictions, it imposes on the agent's rationality, for example, "failing to adjust correctly for repetitions and dependencies in information that they hear multiple times" (Golub and Jackson, 2010, p.113), due to its intuitive appeal and simplicity, the linear combination or the weighted average still serves as "a useful and tractable first approximation" (Golub and Jackson, 2010, p.113) of how a group member pools others' opinions and thereafter forms her own opinion.

The difference between the DeGroot model and the threshold model lies in their representations of the agent's belief. The credence vector is used in the DeGroot model. It is not impractical to assume that each agent assigns her subjective probability distribution to a given proposition. Considering the process of opinion pooling, however, it seems too impractical to require that each agent is explicitly aware of her subjective probability and meanwhile has access to others' subjective probability even if we assume that all the agents fully communicate with each other and have the capability of computing the weighted average. The belief vector used in the threshold model is thus more practical. After all, it is not that hard to answer a yes-or-no question, especially about one's own belief. And it is not that hard either to convey one's own yes-or-no belief to others.

Despite the difference, the views of both models on the updating process are the same. The opinion pooling is executed by each agent consciously or unconsciously round by round. And each execution determines the agents' beliefs in the next round. Thus the process is deterministic.

What if we view the linear combination a little differently? In the model proposed in this section, the linear combination does not determine each agent's belief in the next round but only indicates what each agent would tend to believe in the next round. The updating process determined by the linear combination is thus not deterministic but indeterministic.

In contrast with the DeGroot model, our model deals only with the agent's binary qualitative belief as the threshold model does. In contrast with both the DeGroot model and the threshold model, in our model, the linear combination only determines the probability of the agent's belief change rather than how the agent's belief changes. The linear combination tells us the probability of one belief vector $\mathbf{b}$ updating to another belief vector $\mathbf{b}'$, or equivalently, the probability of the group members being in the doxastic state $\mathbf{b}'$ in the next step, given that the group members are currently in the doxastic state $\mathbf{b}$.

To make the shift of perspective clear, we define the new model formally.

5.3.1. Definition (Indeterministic Binary DeGroot Model). An indeterministic binary DeGroot Model (IBDM) is a structure $\mathcal{ID} = (G, \mathbf{T}, \mathfrak{b})$ where G is a group of m agents and $\mathbf{T}$ is an m-by-m trust matrix. $\mathfrak{b}$ is the set of belief vectors $\{\mathbf{b} \mid \mathbf{b}_i \in \{0, 1\}\}$ where $1 \leq i \leq m$.

Note that each belief vector in $\mathfrak{b}$ represents a possible state of the group members' beliefs about an implicitly given proposition.

5.3.2. EXAMPLE. Given the $\mathbf{T}$ in Example 5.2.1 and $\mathbf{b}$ in Example 5.2.3, the DeGroot model predicts that the next state is

$$\mathbf{c} = \begin{bmatrix} 1 \\ 0.4 \\ 1 \end{bmatrix}.$$

Instead, what the IBDM does is giving each component in $\mathbf{c}$ a new interpretation[4] – the probability of the corresponding agent believing the given proposition in the next step. For example, agent 2 would believe the given proposition in the next step with probability 0.4, according to $\mathbf{c}$. This implies that agent 2 would not believe the given proposition in the next step with probability 0.6. Hence, for each agent, we can compute the probability of believing the given proposition and the probability of not believing the given proposition in this way.

Since each agent updates her belief according to the group members' current beliefs and the updating is independent of other members' updated beliefs, we can compute the probability of the group members' beliefs being in certain state in the next step by multiplying the agents' corresponding probabilities of believing or not believing the given proposition.

For example, the probability of $\mathbf{b}' = [1, 0, 1]^\intercal$ in the next step is $1 \times 0.6 \times 1 = 0.6$. The first number 1 represents the probability of the first agent believing the given proposition in the next step. The second number 0.6 is the probability of the second agent *not* believing the given proposition in the next step. The third number 1 represents the probability of the third agent believing the given proposition in the next step. All these three numbers are given by the credence vector $\mathbf{c}$. Similarly, the probability of $\mathbf{b}'' = [0, 1, 1]^\intercal$ in the next step is $0 \times 0.4 \times 1 = 0$.

The above example shows that for each IBDM, we can compute its corresponding *transition matrix*, as precisely defined below.

Notation. Given a matrix $\mathbf{M}$, we will use $\mathbf{M}_{i*}$ to denote the ith row vector in $\mathbf{M}$ and $\mathbf{M}_{*i}$ to denote the ith column vector in $\mathbf{M}$.

Notation. Let Bv be a random variable for the group's current belief state and Bv' be the random variable for the group's belief state in the next step. Bv_i and Bv'_i denote the random variables for the ith agent's current belief state and her belief state in the next step respectively. $P(Bv'_i = 1 | Bv = \mathbf{b})$ is the probability that agent i would believe the given proposition conditional on the current belief vector $\mathbf{b}$. It can be computed as follows:

$$P(Bv'_i | Bv = \mathbf{b}) = \begin{cases} \mathbf{T}_{i*}\mathbf{b} & Bv'_i = 1 \\ 1 - \mathbf{T}_{i*}\mathbf{b} & Bv'_i = 0 \end{cases} \tag{5.1}$$

[4] The change of interpretation brings some substantial effects, as we will see in the sections to follow and explicitly discuss at the end of Section 5.5

Note that the probability $P(Bv'_i|Bv = \mathbf{b})$ is determined by the linear combination $\mathbf{T}_{i*}\mathbf{b}$ for each agent i in the group. This makes our new perspective precise. In IBDM, we only talk about the probability of transition from one belief state to another belief state rather than the realised and thus deterministic transition from one belief state to another belief state as in the DeGroot model.

5.3.3. Definition. Given an indeterministic binary DeGroot model $\mathcal{ID}$, we define the *transition matrix* of $\mathcal{ID}$ as a 2^m-by-2^m matrix $\mathbb{T}_{\mathcal{ID}}$ with entries

$$\mathbb{T}_{\mathbf{bd}} = P(Bv' = \mathbf{d}|Bv = \mathbf{b}) = \prod_{i \in G} P(Bv'_i = \mathbf{d}_i|Bv = \mathbf{b})$$

where $m = |G|$ and $\mathbf{b}, \mathbf{d} \in \mathfrak{b}$.

To compute the transition matrix, we only need the trust matrix. Once the number of agents in the group G is given, all the possible belief vectors of the group are fixed. The number of the rows/columns of the trust matrix tells us how many members the given group G has. So the trust matrix encodes all the information we need to compute its corresponding transition matrix, as illustrated by the following example.

5.3.4. Example. Given a trust matrix

$$\mathbf{T} = \begin{bmatrix} 0 & 1 & 0 \\ 0.4 & 0 & 0.6 \\ 1 & 0 & 0 \end{bmatrix}$$

by Definition 5.3.3, the transition matrix is

$$\mathbb{T} = \begin{array}{c} \\ 111 \\ 110 \\ 101 \\ 100 \\ 011 \\ 010 \\ 001 \\ 000 \end{array} \begin{array}{c} \begin{array}{cccccccc} 111 & 110 & 101 & 100 & 011 & 010 & 001 & 000 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.4 & 0 & 0.6 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.4 & 0 & 0.6 & 0 \\ 0 & 0.6 & 0 & 0.4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.6 & 0 & 0.4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

The entries in the second row of the transition matrix, i.e. row 110, represent the probabilities for all possible states after one step of transition, given that the initial state is 110 (for brevity, we write the belief vector in the form of a string). Observe that each row of the transition matrix sums up to one.

If we take the square of $\mathbb{T}$, we get a new matrix $\mathbb{T}^2$. The entries in the second row of this new matrix represent the probabilities of all possible states two steps after the initial state 110 respectively.

More generally, for each m-by-m trust matrix, its corresponding transition matrix is a *probability matrix*, that is, $\mathbb{T}_{ij} \in [0,1]$ and $\sum_{j=1}^{2^m} \mathbb{T}_{ij} = 1$ for all $i, j \in \mathbb{N}$ such that $0 < i, j \leq 2^m$. The ijth entry in the matrix $\mathbb{T}^n$ represents the probability of the group members' beliefs being in the jth state after n steps, given the initial state is the ith state. Given a process running from the initial state to the latest state, according to the way we compute the transition matrix from the trust matrix, the process is a Markov chain, because the probability of the process moving from state i to state j with one step does not depend on any other states before state i in the process.

Therefore, any questions about the transition matrix for a discrete Markov chain can also be asked about the transition matrix in the IBDM. Of them, the first question we ask is whether the powers of the transition matrix in the IBDM converge, or under what conditions it converges. The convergence means that the transition probability tends to get stable after large enough steps and it thus conceptually implies that there is a tendency for the group to have a certain belief state.

5.4 The Convergence of the Transition Matrix

In this section, we prove a sufficient condition for the powers of the transition matrix in the IBDM to be convergent by revealing a connection between the trust matrix and its corresponding transition matrix in the IBDM.

We start with an observation about the transition matrix in the IBDM.

Notation. We write $\mathbf{1}$ (resp. $\mathbf{0}$) for the constant vector with all entries equal to 1 (resp. 0).

It is not hard to realise that no matter what the trust matrix is, $\mathbb{T}_{\mathbf{11}}$ and $\mathbb{T}_{\mathbf{00}}$ are always 1. It means that if the group has reached a consensus, then it will keep the same consensus and not change anymore. The states of this kind are called "absorbing states" in a Markov chain. So the two consensus states $\mathbb{T}_{\mathbf{11}}$ and $\mathbb{T}_{\mathbf{00}}$ are both absorbing states.

5.4.1. Definition. A state s_i of a Markov chain is called absorbing if it is impossible to leave it ($\mathbb{T}_{ii} = 1$).

It does not mean much that there are absorbing states in a Markov chain. However, if we can ensure that it is possible to go to at least one absorbing state (not necessarily in one step) from every state in the Markov chain, then it makes a lot of difference.

5.4.2. Definition. A Markov chain is absorbing if there exists in it at least one absorbing state, and if for every state the probability of reaching an absorbing state (not necessarily in one step) is strictly positive.

5.4.3. Theorem (Theorem 11.3 in Grinstead and Snell (1997)). *In an absorbing Markov chain, the probability that the process will be absorbed is 1. That is, for every state i in the absorbing Markov chain,*

$$\sum_{s \text{ is absorbing}} \mathbf{P}^{\infty}_{is} = 1$$

where $\mathbf{P}$ is the transition matrix for the Markov chain and $\mathbf{P}^{\infty} = \lim_{n\to\infty} \mathbf{P}^n$.

So if the transition matrix in the IBDM is a transition matrix for an absorbing Markov chain, then no matter what the initial state of the group members' beliefs is, it will finally reach an absorbing state with some probability and stay there afterwards. We know from before that $\mathbf{1}$ and $\mathbf{0}$ are two absorbing states. It means that it is *possible* for the group to finally reach a consensus and never disagree again if the Markov process is absorbing. Next, we will show that some properties of the *trust matrix* can ensure the transition matrix to be absorbing. Moreover, these properties also ensure that the group would finally reach a consensus.

First, notice the fact that for every trust matrix $\mathbf{T}$, we can define its associated graph.

5.4.4. Definition. Given an IBDM where the group of agent is G and the trust matrix is $\mathbf{T}$, its associated graph is $\mathcal{SN}_{\mathbf{T}} = (G, E)$ whose nodes are given by the group G and whose edges in E are given by the following condition: $(i, j) \in E$ if and only if $\mathbf{T}_{ij} > 0$ where $i, j \in G$.

Note that given the trust matrix's associated graph $\mathcal{SN}_{\mathbf{T}}$, $\mathbf{T}^n_{ij} > 0$ if and only if there is an n-step walk from i to j. An *n-step walk* in $\mathcal{SN}_{\mathbf{T}}$ from i to j is a sequence of links $(i_0, i_1), \ldots, (i_{n-1}, i_n)$ such that $(i_k, i_{k+1}) \in E$ for each $k \in \{0, \ldots, n-1\}$, with $i_0 = i$ and $i_n = j$. When all nodes in $\{i_0, \ldots, i_n\}$ are distinct from each other with the possible exception of $i_0 = i_n$, the n-step walk from i to j is an *n-step path* from i to j.

Given all these preliminaries, we are ready to define the properties of the trust matrix, or equivalently, the properties of the trust matrix's associated graph – strong connectedness and aperiodicity.

5.4.5. Definition (Strongly connected probability matrix). A graph is strongly connected if all nodes can reach each other via a path. Given a probability matrix $\mathbf{T}$, we say that $\mathbf{T}$ is *strongly connected* if and only if its associated graph $\mathcal{SN}_{\mathbf{T}}$ is strongly connected.

If we take the graph as the group's social network, the strong connectedness is not a very implausible property to impose on the social network, especially when considering the small world phenomenon (Watts and Strogatz, 1998; Travers and Milgram, 1969). We will call the group in the IBDM whose trust matrix is strongly connected a "*community*".

The following theorem tells us that for a community G in an IBDM, in every Markov chain which consists of states from $\mathfrak{b}$ of the IBDM, $\mathbf{1}$ and $\mathbf{0}$ are the *only* two absorbing states.

5.4.6. THEOREM (Theorem 11.10 in Grinstead and Snell (1997)). *For a strongly connected trust matrix* $\mathbf{T}$*, every column vector* $\mathbf{x}$ *such that* $\mathbf{Tx} = \mathbf{x}$ *is a constant vector.*

By noticing that for every IBDM a state $\mathbf{b} \in \mathfrak{b}$ is absorbing according to the transition matrix if and only if $\mathbf{Tb} = \mathbf{b}$, the result follows:

5.4.7. COROLLARY. *For every IBDM whose trust matrix is strongly connected, there exist two and only two absorbing states in the transition matrix of the IBDM. That is,* $\boldsymbol{1}$ *and* $\boldsymbol{0}$*.*

Although the strongly connected trust matrix in the IBDM ensures the existence of two and only two absorbing states according to its transition matrix, it cannot ensure that the Markov chain according to the transition matrix is absorbing. It is possible that some states cannot go to one of the two absorbing states. The following strongly connected trust matrix exemplifies this possibility.

5.4.8. EXAMPLE. Given a trust matrix

$$\mathbf{T} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{5.2}$$

its corresponding transition matrix is

$$\mathbb{T} = \begin{array}{c} \\ 11 \\ 10 \\ 01 \\ 00 \end{array} \begin{array}{c} \begin{array}{cccc} 11 & 10 & 01 & 00 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array} \tag{5.3}$$

By computing $\mathbb{T}^n$ for several values of n, we find the following pattern. $\mathbb{T}^n = \mathbb{T}$ when n is odd; When n is even, $\mathbb{T}^n$ is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The Markov chain according to the transition matrix 5.3 in the above example is not absorbing, because 10 only goes to 01 and 01 only goes to 10, neither of which is an absorbing state. It also causes the oscillation of $\mathbb{T}^n$ between two different matrices as n tends to infinity. Although the phenomenon of oscillation is interesting in itself and worth an extensive study, we will not touch on this topic here and refer the readers to van Benthem (2015*b*), which studies the oscillation from the perspective of logic. Instead, we try to find out how we can avoid the oscillation.

The reason for the powers of transition matrix to oscillate in Example 5.4.8 is that no agent in the group trusts himself. It indicates that such oscillation may be avoided by requiring that at least one group member trusts herself to at least some (positive) degree. This sounds a very weak condition and not impractical. After all, it is hard to imagine all people in the society are just following others.

The condition that at least one group member trusts herself, i.e. $\mathbf{T}_{ii} > 0$ for some $i \in G$, indeed works. Even a weaker condition – aperiodicity – on the trust matrix[5], together with the condition of strong connectedness, can ensure that its corresponding transition matrix gives rise to an absorbing Markov chain.

5.4.9. Definition. [6] Given a trust matrix $\mathbf{T}$, the period of an agent i in the model is the greatest common divisor of the members in the set $\{n \in \mathbb{N} \mid \mathbf{T}^n_{ii} > 0\}$:

$$g(i) = \gcd\{n \in \mathbb{N} \mid \mathbf{T}^n_{ii} > 0\} \ .$$

i is *aperiodic* if $g(i) = 1$ and periodic if $g(i) > 1$. $\mathbf{T}$ is aperiodic if and only if all the agents in it are aperiodic.

5.4.10. Lemma. *Given an indeterministic binary DeGroot model $\mathcal{ID}$, if the trust matrix in the IBDM is strongly connected and aperiodic, then given any vector $\mathbf{b} \in \mathfrak{b}$ with some $i \in G$ such that $\mathbf{b}_i = 1$, there exists a path from $\mathbf{b}$ to $\mathbf{1}$. I.e. there exists a sequence $p^0, p^1, \ldots, p^n$ of vectors in $\mathfrak{b}$ such that $p^0 = \mathbf{b}$, $p^n = \mathbf{1}$ and for all natural numbers $k \in [0, n-1]$, $\mathbb{T}_{p^k, p^{k+1}} > 0$.*

Proof:
The proof can be found in Appendix D.3. □

5.4.11. Theorem. *Given an indeterministic binary DeGroot model $\mathcal{ID}$, if the trust matrix in the IBDM is strongly connected and aperiodic, then the transition matrix for the IBDM is a transition matrix for an absorbing Markov chain including only two absorbing states.*

---

[5] Take the trust matrix in Example 5.2.1. It is aperiodic but there is no $i \in G$ such that $\mathbf{T}_{ii} > 0$

[6] The definition of aperiodicity we present here is different from another definition of aperiodicity appearing in the literature, for example, Jackson (2010, Chapter 8). We prove a result in Appendix D.1 stating that these two definitions are equivalent given that the trust matrix is strongly connected.

Proof:
The proof follows from the last lemma and the definition of the absorbing Markov chain and Corollary 5.4.7. □

Together with Theorem 5.4.3, the following result follows immediately.

5.4.12. Corollary. *Given an indeterministic binary DeGroot model $\mathcal{ID}$, if the trust matrix in the IBDM is strongly connected and aperiodic, then the powers of the transition matrix for the IBDM converge to a limiting matrix.*

In Appendix D.4, we briefly sketch how Theorem 5.4.3 is proved (which will help understand Corollary 5.4.12 better) and show the benefits brought by shifting the perspective on the DeGroot model.

Next, we introduce the notion of potential group belief, to which the convergence result we have just established will be key once we restrict our attention to the group with a strongly connected and aperiodic trust matrix.

5.5 Group's Potential Belief

As we indicated at the very beginning of this Chapter, instead of a fierce way of resolving the conflict between beliefs of different group members, a softer way is pursued. It puts more focus on the tendency of how the group members' beliefs evolve, which is driven by the group members' mutual influence through its social network. The group's belief is then decided by the tendency of the evolving of the group members' beliefs.

5.5.1. Definition (Group's potential belief). Given an IBDM $\mathcal{ID}$ and the group G's initial belief state $\mathbf{b}$, the group G in $\mathcal{ID}$ tends to believe the given proposition if and only if there exists a natural number N such that for all $n \geq N$,

$$\mathbb{T}^n_{\mathbf{b1}} > 0.5\,.$$

The definition says that the group has potential belief of a given proposition if and only if after several steps, say N steps, it is always very probable (in the sense of being larger than 0.5) that the group would reach the positive consensus on the given proposition. This definition thus features three points: social interaction, high probability of the tendency to the consensus ($\mathbf{1}$) and stability. The social interaction is simplified in the IBDM as a trust matrix and an iteration of updating by weighted average. The probability given by the IBDM is the probability of transition from one belief state to another belief state. The definition thus bases the notion of group's potential belief on the objective transition probability. It makes use of the 0.5-threshold rule, just like the Lockean thesis for the single agent's belief (Foley, 1992). The difference is that for the single agent the

Lockean thesis relies on the agent's subjective probability and for the group the potential belief relies on the objective transition probability. The last feature of the definition is that it requires the persistence of high transition probability to the positive consensus after N steps in the updating process. This persistence or the stability makes sure that the notion of potential belief is not a contingent artefact.

The three nice features of the definition come with a difficult decision problem – how to efficiently decide whether the group has the potential belief given an IBDM. We have no general solution which can be applied to all types of groups. Nevertheless, to the group whose trust matrix is strongly connected and aperiodic, we know how to decide whether the group can finally reach the potential belief. It is not hard to understand why this is the case by recalling our result in the previous section. That is, the powers of the transition matrix for the group with a strongly connected and aperiodic trust matrix always converges to a matrix. With the convergence result, the definition of potential belief for a group whose trust matrix is strongly connected and aperiodic can be simplified.

5.5.2. PROPOSITION. *Given an IBDM $\mathcal{ID}$, if the trust matrix is strongly connected and aperiodic, and the group G's initial belief state is $\mathbf{b}$, then the group G tends to believe the given proposition if and only if*

$$\mathbb{T}^{\infty}_{\mathbf{b1}} > \mathbb{T}^{\infty}_{\mathbf{b0}}$$

where $\mathbb{T}$ is the transition matrix for the IBDM.

Proof:
Let $\mathbb{T}$ be the transition matrix for the IBDM. By Theorem 5.4.11 and Theorem 5.4.3 and the assumption that the trust matrix is strongly connected and aperiodic, $\mathbb{T}^n$ converges as n tends to ∞, so $\mathbb{T}^\infty$ exists. Moreover, by the same theorem, we know that $\mathbb{T}$ is a transition matrix for an absorbing Markov chain with only two absorbing states. So given any initial belief vector, the probability is only distributed between $\mathbf{1}$ and $\mathbf{0}$ in $\mathbb{T}^\infty$. This implies that $\mathbb{T}^{\infty}_{\mathbf{b1}} > \mathbb{T}^{\infty}_{\mathbf{b0}}$ if and only if $\mathbb{T}^{\infty}_{\mathbf{b1}} > 0.5$. □

Next, we show how to compute the entries in the limiting matrix of the powers of the transition matrix ($\mathbb{T}^\infty$) generated by a strongly connected and aperiodic trust matrix. To this end, a more precise characterisation of the relationship between the trust matrix and its generated transition matrix is needed. The following theorem fulfils this task. It says that given the group's current belief state $\mathbf{b}$, the probability of agent i changing her belief state to 1 after n steps is distributed among those group belief states where agent i's belief state is 1 after n steps.

5.5.3. THEOREM. *Given an indeterministic binary DeGroot model $\mathcal{ID}$, for all* $\mathbf{b}, \mathbf{s} \in \mathfrak{b}$*,* $i \in G$ *and* $n \in \mathbb{N}$*,*

$$\mathbf{T}^n_{i*} \cdot \mathbf{b} = \sum_{\mathbf{s}_i = 1} \mathbb{T}^n_{\mathbf{bs}}$$

where $\mathbf{T}_{i*}$ *is the ith row of the trust matrix* $\mathbf{T}$ *and* $\mathbf{b}$ *is the group's current belief state.*

Proof:
The proof can be found in Section D.2.2. □

5.5.4. EXAMPLE. To illustrate this theorem, take the trust matrix in Example 5.3.4, let $\mathbf{b} = 110$, then $\mathbf{T}_{2*}\mathbf{b} = 0.4$. And $\sum_{\mathbf{s_2}=1} \mathbb{T}_{\mathbf{bs}} = 0.4$, as expected. We leave it to the readers to check those cases where $n > 1$ for the same example.

Note that the above theorem holds for all trust matrices and its generated transition matrices. An immediate corollary of the above theorem is the following one.

5.5.5. COROLLARY. *Given an IBDM* $\mathcal{ID}$*, as n tends to* ∞*, if the powers of its transition matrix* $\mathbb{T}^n$ *converge, then the powers of the trust matrix* $\mathbf{T}^n$ *converge too.*

Proof:
Assume that as n tends to ∞ the IBDM's transition matrix $\mathbb{T}^n$ converges. By theorem 5.5.3,

$$\mathbf{T}^n_{ij} = \mathbf{T}^n_{i*}\mathbf{e_j} = \sum_{\mathbf{s}_i = 1} \mathbb{T}^n_{\mathbf{e}_j\mathbf{s}}$$

for all $n \in \mathbb{N}$, where $\mathbf{e_j}$ is the vector where the ith entry is 1 and the other entries are all 0s. Because $\mathbb{T}^n$ converges to $\mathbb{T}^\infty$, $\mathbf{T}^n_{ij}$ converges to

$$\sum_{\mathbf{s}_i = 1} \mathbb{T}^\infty_{\mathbf{e}_j\mathbf{s}} \quad .$$

Each entry of $\mathbf{T}$ thus converges to a number. Therefore, the trust matrix $\mathbf{T}$ converges. □

Together with Theorem 5.4.3, the corollary tells us that if the trust matrix is strongly connected and aperiodic, then $\mathbf{T}^\infty$ exists. Moreover, together with some other observations, it tells us how $\mathbf{T}^\infty$ looks like.

5.5.6. THEOREM. *Given an indeterministic binary DeGroot model* $\mathcal{ID}$*, if the trust matrix* $\mathbf{T}$ *is strongly connected and aperiodic, then as* $n \to \infty$*, the powers* $\mathbf{T}^n$ *approach a limiting matrix* $\mathbf{W}$ *with all rows the same vector* $\mathbf{w}$*. The vector* $\mathbf{w}$ *is a strictly positive probability vector (i.e. the components are all positive and sum to one).*

Proof:
The proof of this theorem is essentially the same as what we spell out in Appendix D.2.3.

What is worth noticing is that this theorem is one step away from the following classical convergence result for Markov chain.

5.5.7. THEOREM (Theorem 11.7 in Grinstead and Snell (1997)). *Let* $\mathbf{P}$ *be the transition matrix for a* regular *Markov chain. Then as* $n \to \infty$*, the powers* $\mathbf{P}^n$ *approach a limiting matrix* $\mathbf{W}$ *with all rows the same vector* $\mathbf{w}$*. The vector* $\mathbf{w}$ *is a strictly positive probability vector (i.e. the components are all positive and sum to one).*

The missing step is the proof of the following two statements:

- the transition matrix for a regular Markov chain must be strongly connected and aperiodic; and
- the strongly connected and aperiodic trust matrix must be a transition matrix for a regular Markov chain.

We fill the gap in Appendix D.3 by proving Theorem D.3.1.

Note that the way we prove Theorem 5.5.7 (the whole proof can be found in Appendix D.2) is different from the standard proof and Doeblin's proof, both of which can be found in Grinstead and Snell (1997, Section 11.4). □

Given the above theorem, we can have the following corollary, which opens the door towards the computation of $\mathbb{T}^\infty$.

5.5.8. COROLLARY. *Given an indeterministic binary DeGroot model* $\mathcal{ID}$*, if the trust matrix is strongly connected and aperiodic, then for all* $\mathbf{b}, \mathbf{s} \in \mathfrak{b}$ *and* $i \in G$*,*

$$\mathbf{T}^\infty_{i*} \cdot \mathbf{b} = \sum_{s_i=1} \mathbb{T}^\infty_{\mathbf{bs}} = \mathbb{T}^\infty_{\mathbf{b1}}$$

where $\mathbf{T}^\infty$ *is a matrix with all rows the same vector* $\mathbf{w}$*. The vector* $\mathbf{w}$ *is a strictly positive probability vector (i.e., the components are all positive and they sum to one).*

This corollary provides a way of computing $\mathbb{T}^\infty_{\mathbf{b1}}$. We only need to know the row vector $\mathbf{w}$ in $\mathbf{T}^\infty$, which turns out to be in the left nullspace of the matrix $\mathbf{T} - \mathbf{I}$ because $\mathbf{wT} = \mathbf{wI}$. We can thus first compute the left nullspace of $\mathbf{T} - \mathbf{I}$, and then normalising a vector in the nullspace to get the probability vector $\mathbf{w}$. Therefore, we have achieved an algorithm for computing all the entries in the limiting matrix $\mathbb{T}^\infty$. And we can decide whether the group tends to believe the given proposition by comparing $\mathbb{T}^\infty_{\mathbf{b1}}$ and $\mathbb{T}^\infty_{\mathbf{b0}}$ given that the initial state is $\mathbf{b}$.

At last, we round up our analysis of the relationship between the trust matrix and its generated transition matrix by presenting the following theorem:

5.5.9. THEOREM. *Given an IBDM $\mathcal{ID}$, if the trust matrix* $\mathbf{T}$ *is strongly connected, then the following four statements are equivalent:*

1. *the **trust** matrix* $\mathbf{T}$ *is **aperiodic**;*
2. *the **transition** matrix generated from* $\mathbf{T}$ *is a transition matrix for an **absorbing** Markov chain;*
3. *the powers of the **transition** matrix generated from* $\mathbf{T}$ *converge;*
4. *the powers of the **trust** matrix* $\mathbf{T}$ *converge.*

Proof:
We have proved the direction from 1 to 2 (Theorem 5.4.11) and from 3 to 4 (Corollary 5.5.5). The direction from 2 to 3 is by Theorem 5.4.3. The direction from 4 to 1 is by LEMMA 4 in Appendix A of Golub and Jackson (2010). □

The above theorem may give readers with the impression that there is no need of shifting the perspective and taking the indeterministic interpretation. After all, when considering the strongly connected and aperiodic trust matrix, Corollary 5.5.8 and Proposition 5.5.2 tell us that we can have a mathematically equivalent definition of the group's potential belief even if we stick to the DeGroot model and its deterministic perspective.

5.5.10. PROPOSITION. *Given an IBDM $\mathcal{ID}$, if the trust matrix is strongly connected and aperiodic and the group G's initial belief state* $\mathbf{b}$*, then the group G tends to believe the given proposition if and only if*

$$\mathbf{T}^{\infty}_{i*}\mathbf{b} > 0.5$$

where $\mathbf{T}$ *is the trust matrix for the IBDM.*

Then what is the added value of shifting the perspective and taking the indeterministic interpretation?

Notice that the observation given by the above proposition about the equivalent definition of potential group belief does not apply to the general case. For Definition 5.5.1, there is no way of even having a mathematically equivalent definition in the DeGroot model, because there is no way of extracting the number $\mathbb{T}^n_{\mathbf{b1}}$ from the DeGroot model. As Theorem 5.5.3 shows, what the DeGroot model tells us about the transition matrix is just $\sum_{\mathbf{s}_i=1}\mathbb{T}^n_{\mathbf{bs}}$, but not how this number is distributed among all those belief vectors in the set $\{\mathbf{s} \in \mathfrak{b} \mid \mathbf{s}_i = 1\}$.

Even if we consider only the strongly connected and aperiodic trust matrix, the shifting of perspective makes substantial difference. Firstly, in the IBDM, when the transition matrix is for an absorbing Markov chain, by computing its fundamental matrix (c.f. Appendix D.4), we can know a lot of information about the absorbing Markov chain. For example, given an initial belief state $\mathbf{b}$ of the group, the expected number of times the chain is in belief state $\mathbf{d}$; and given that the chain starts in the group's belief state $\mathbf{b}$, the expected number of steps before the chain is absorbed. However, in a DeGroot model, there is nowhere to find a Markov chain. Secondly, the IBDM restricts the group's belief states to a space where only binary vectors are allowed. This does not only justify the necessity of extracting a qualitative notion of group belief from the probability, but also paves the way for studying the logic of the notion of the group's potential belief.

5.6 Kripke-DeGroot Frame

From this section on, we pursue the question concerning the logic of potential group belief: whether the group's potential belief is consistent, whether it is closed under conjunction and how we can systematically decide the validity of a statement about potential belief. To understand these questions, the IBDM is not rich enough, because the IBDM only deals with a single proposition (implicitly given), while logic is about the relationship between multiple propositions. This is why we introduce Kripke to DeGroot.

5.6.1. Definition. A Kripke-DeGroot frame (KDF) is a structure

$$\mathcal{KD} = (W, R_i, \mathbf{T})$$

where

- (W, R_i) is a Kripke frame in Definition 1.2.1 with its accessibility relation indexed with the label for each group member in an implicitly given group G;
- $\mathbf{T}$ is a trust matrix for the given group of agents G;

The set of possible worlds is introduced to equip the model with the ability to express propositions explicitly. Each subset of W is taken as a proposition, and the set of all subsets of W constitutes a Boolean algebra. The Kripke-DeGroot frame can thus specify each agent's current belief state for each proposition. $R_i(w) \subseteq Q$ means that agent i believes Q in the world w.

In this section, we assume that R_i is serial for all $i \in G$ (for all $w \in W$, there is $v \in W$ such that $R_i wv$), which assures that each group member's belief is fully consistent. Moreover, for the purpose of illustration, we simplify the setting by requiring that given a Krikpke-DeGroot frame, for all $w, v \in W$ and for all $i \in G$, $R_i(w) = R_i(v)$. Whenever this uniformity of R_i is assumed, we will use $f(i)$ to denote $R_i(w)$ since w as a parameter does not play a role any more. This will save us from an overdose of indexes attached to the notations in the following analysis.

For different propositions are explicitly modeled in the Kripke-DeGroot frame, we can represent the group's belief state for each proposition by a binary vector decided by each agent's belief state for that proposition.

5.6.2. DEFINITION (Group's belief state for Q). Given a KDF and any $Q \subseteq W$, the group's current belief state is specified by $\mathbf{b}^Q$, a binary vector satisfying

$$\mathbf{b}_i^Q = \begin{cases} 1 & f(i) \subseteq Q \\ 0 & \text{otherwise} \end{cases} \tag{5.4}$$

Notation. Let $\overline{Q} = W - Q$ be the negation of proposition Q and $\overline{\mathbf{b}} = \mathbf{1} - \mathbf{b}$.

The relation between different propositions decides the relation between the group's belief states with respect to the different propositions.

5.6.3. PROPOSITION. *Given a KDF,*

a For all $S, Q \subseteq W$, if $S \subseteq Q$, then $\mathbf{b}^S \leq \mathbf{b}^Q$, i.e. for all $i \in G$, $\mathbf{b}_i^S \leq \mathbf{b}_i^Q$

b If R_i is serial for all $i \in G$, then for all $Q \subseteq W$, $\mathbf{b}^{\overline{Q}} \leq \overline{\mathbf{b}^Q}$

Note that Proposition 5.6.3.b relies on the assumption that R_i is serial for all $i \in G$.

Since a KDF specifies the group's current belief state for all $Q \subseteq W$ we have the definition of the group's potential belief as follows.

5.6.4. DEFINITION (Group's potential belief of Q). Given an KDF $\mathcal{KD}$, the group G in $\mathcal{KD}$ tends to believe Q if and only if there exists a natural number N such that for all $n \geq N$,

$$\mathbb{T}^n_{\mathbf{b}^Q \mathbf{1}} > 0.5$$

where $\mathbb{T}$ is the transition matrix generated from the trust matrix in $\mathcal{KD}$, $\mathbf{b}^Q$ is the vector representing each agent's belief about Q and $\mathbf{1}$ is the constant vector with each entry equal to 1.

Within the Kripke-DeGroot frame, we can address the questions concerning the logic of the group's potential belief. Note that our discussion will be restricted to the class of frames whose trust matrix is strongly connected and aperiodic. The results we achieve about the logic rely on the convergence of the transition matrix generated from the strongly connected and aperiodic trust matrix. Although the results are restricted, they are not of limited value, considering that strongly connectedness and aperiodicity are both very common properties of social networks. We have mentioned that strongly connectedness is confirmed to be a pervasive property in many networks by a considerable amount of studies (Travers and Milgram, 1969; Kautz, Selman and Shah, 1997; Watts and Strogatz, 1998; Milgram, 1967). As for aperiodicity, it is not a rare property either although its definition appears to be complicated. Notice that reflexivity (for all $i \in G$, $\mathbf{T}_{ii} > 0$) is a special case of aperiodic. Moreover, when the trust matrix is strongly connected, even a single reflexive agent (there exists $i \in G$ such that $\mathbf{T}_{ii} > 0$) can imply the aperiodicity of the trust matrix. It is not impractical to expect that there is at least one person in the group having at least a little trust in herself.

Recall that we call a group with a strongly connected trust matrix a "community". Subsequently, we will call the group with a strongly connected and aperiodic trust matrix a "*regular community*" and the KDF with a strongly connected and aperiodic trust matrix a "*regular KDF*".

In the remaining part of this section, we approach two specific questions on the logical properties of the regular community's potential belief to nurture readers' understanding. In the next section, we construct a logic which subsumes the logic of the regular community's potential belief.

Closure upwards

The first question is whether the regular community tends to believe Q if it tends to believe a stronger proposition $S \subseteq Q$.

To answer the first question, recall Proposition 5.5.2 and Theorem 5.5.3.

5.6.5. PROPOSITION. *Given a regular KDF $\mathcal{KD}$, for all $S, Q \subseteq W$ such that $S \subseteq Q$, if the regular community tends to believe S, then it tends to believe Q.*

Proof:
Assume that $S \subseteq Q$ and that the given KDF is regular. By Proposition 5.6.3.a, $\mathbf{b}^S \leq \mathbf{b}^Q$. Hence for all $i \in G$ and $n \in \mathbb{N}$, $\mathbf{T}^n_{i*}\mathbf{b}^S \leq \mathbf{T}^n_{i*}\mathbf{b}^Q$. By Theorem 5.5.3 and Corollary 5.5.8, $\mathbb{T}^\infty_{\mathbf{b}^S\mathbf{1}} \leq \mathbb{T}^\infty_{\mathbf{b}^Q\mathbf{1}}$. By Proposition 5.5.2 and the assumption that the regular community tends to believes S,

$$0.5 < \mathbb{T}^\infty_{\mathbf{b}^S\mathbf{1}} \leq \mathbb{T}^\infty_{\mathbf{b}^Q\mathbf{1}} \ ,$$

which in turn implies that the regular community tends to believe Q. □

Mutual consistency

The second question is whether it is possible for the regular community to have potential belief of both Q and $\overline{Q}$.

5.6.6. LEMMA. *Given a KDF $\mathcal{KD}$, for all $\mathbf{b}, \mathbf{s} \in \mathfrak{b}$ and $n \in \mathbb{N}$,*

$$\mathbb{T}^n_{\mathbf{bs}} = \mathbb{T}^n_{\overline{\mathbf{b}}\overline{\mathbf{s}}} \ .$$

Proof:
We prove $\mathbb{T}^n_{\mathbf{bs}} = \mathbb{T}^n_{\overline{\mathbf{b}}\overline{\mathbf{s}}}$ by induction.

First, we show that for all $x \in G$, $P(Bv'_x = 1|Bv = \overline{\mathbf{b}}) = P(Bv'_x = 0|Bv = \mathbf{b})$, which will implies that $\mathbb{T}_{\mathbf{bs}} = \mathbb{T}_{\overline{\mathbf{b}}\overline{\mathbf{s}}}$.

$$\begin{aligned}
P(Bv'_x = 1|Bv = \overline{\mathbf{b}}) &= \sum_{i \in G} \mathbf{T}_{xi} \overline{\mathbf{b}}_i && (5.5)\\
&= \sum_{i \in G} \mathbf{T}_{xi}(1 - \mathbf{b}_i) && (5.6)\\
&= \sum_{i \in G} \mathbf{T}_{xi} - \sum_{i \in G} \mathbf{T}_{xi}\mathbf{b}_i && (5.7)\\
&= 1 - P(Bv'_x = 1|Bv = \mathbf{b}) && (5.8)\\
&= P(Bv'_x = 0|Bv = \mathbf{b}) && (5.9)
\end{aligned}$$

Hence it follows that for all $x \in G$, $P(Bv'_x = \mathbf{s}_x|Bv = \mathbf{b}) = P(Bv'_x = \overline{\mathbf{s}}_x|Bv = \overline{\mathbf{b}})$, which in turn implies that $\mathbb{T}_{\mathbf{bs}} = \prod_{x \in G} P(Bv'_x = \mathbf{s}_x|Bv = \mathbf{b}) = \prod_{x \in G} P(Bv'_x = \overline{\mathbf{s}}_x|Bv = \overline{\mathbf{b}}) = \mathbb{T}_{\overline{\mathbf{b}}\overline{\mathbf{s}}}$.

Second, given $\mathbb{T}^n_{\mathbf{bs}} = \mathbb{T}^n_{\overline{\mathbf{b}}\overline{\mathbf{s}}}$, we show that $\mathbb{T}^{n+1}_{\mathbf{bs}} = \mathbb{T}^{n+1}_{\overline{\mathbf{b}}\overline{\mathbf{s}}}$

$$\begin{aligned}
\mathbb{T}^{n+1}_{\mathbf{bs}} &= \sum_{\mathbf{k} \in \mathfrak{b}} \mathbb{T}^n_{\mathbf{bk}} \mathbb{T}_{\mathbf{ks}} && (5.10)\\
&= \sum_{\mathbf{k} \in \mathfrak{b}} \mathbb{T}^n_{\overline{\mathbf{b}}\overline{\mathbf{k}}} \mathbb{T}_{\overline{\mathbf{k}}\overline{\mathbf{s}}} && (5.11)\\
&= \mathbb{T}^{n+1}_{\overline{\mathbf{b}}\overline{\mathbf{s}}} && (5.12)
\end{aligned}$$

□

With the lemma, we can answer the second question.

5.6.7. PROPOSITION. *Given a regular KDF $\mathcal{KD}$ in which R_i is serial for all $i \in G$, if the regular community tends to believe Q, then it does not tend to believe $\overline{Q}$.*

Proof:
Assume that the regular community in $\mathcal{KD}$ tends to believe Q, which by Proposition 5.5.2 implies that $\mathbb{T}^\infty_{\mathbf{b}^Q\mathbf{1}} > \mathbb{T}^\infty_{\mathbf{b}^Q\mathbf{0}}$. Because $\mathbb{T}^\infty_{\mathbf{b}^Q\mathbf{1}} + \mathbb{T}^\infty_{\mathbf{b}^Q\mathbf{0}} = 1$ by Theorem 5.4.11 and Corollary 5.4.12, $\mathbb{T}^\infty_{\mathbf{b}^Q\mathbf{0}} < 0.5$. By Lemma 5.6.6,

$$\mathbb{T}^\infty_{\overline{\mathbf{b}^Q}\mathbf{1}} = \mathbb{T}^n_{\mathbf{b}^Q\mathbf{0}} < 0.5 \tag{5.13}$$

By Proposition 5.6.3.b, $\overline{\mathbf{b}^Q} \geq \mathbf{b}^{\overline{Q}}$. Together with Theorem 5.5.3, it implies that

$$\mathbb{T}^\infty_{\mathbf{b}^{\overline{Q}}\mathbf{1}} \leq \mathbb{T}^\infty_{\overline{\mathbf{b}^Q}\mathbf{1}} \quad .$$

Because of equation 5.13,

$$\mathbb{T}^\infty_{\mathbf{b}^{\overline{Q}}\mathbf{1}} < 0.5$$

which implies that the regular community does not tend to believe $\overline{Q}$. This completes the proof. □

The answers to the two questions together give us the following result about the consistency of the regular community's potential belief.

5.6.8. Corollary. *Given a regular KDF $\mathcal{KD}$ in which for all $i \in G$ R_i is serial, for all $S, T \subseteq W$ such that $S \cap T = \varnothing$, if the regular community tends to believe S, then it does not tend to believe T, .*

Note that this corollary only says that the notion of the regular community's potential belief does not allow mutual inconsistency. But the regular community may tend to believe Q_1, believe Q_2, believe Q_3, ... and believe Q_n with $n \geq 3$ while $\bigcap_i Q_i = \varnothing$.

What else logical rules should the regular community's potential belief obey? To systematically solve this problem, we study the logic of the regular community's potential belief in the next section.

5.7 Logic of the Regular Community's Potential Belief

In this section, we propose a logic where potential group belief can be expressed. It helps understand potential group belief better conceptually and logically.

Before going into the details of the logic, we recall an observation we made at the end of Section 5.5 about the definition of the regular community's potential belief. This observation has been implicitly used in our two answers (Proposition 5.6.5 and 5.6.7) in the last section.

5.7.1. Proposition. *Given a regular KDF, the regular community tends to believe Q if and only if*

$$\mathbf{T}^{\infty}_{i*}\mathbf{b}^{Q} > 0.5$$

As we have remarked at the end of Section 5.5, this way of seeing the regular community's potential belief is conceptually different. However, with respect to the regular community, it is mathematically equivalent and simplifies the computation as we have shown in Section 5.5. Hence we approach the logic of the regular community's potential belief via this conceptually different way.

Then what is the conceptual interpretation of the limiting matrix $\mathbf{T}^{\infty}$ of the regular community's trust matrix? Each entry of the trust matrix $\mathbf{T}_{ij}$ tells us how much influence agent j can have on agent i directly. More generally, $\mathbf{T}^{n}_{ij}$ tells us how much influence agent j can have on agent i through those paths from j to i with n steps. So $\mathbf{T}^{\infty}_{ij}$ means the influence agent j can have on agent i by running the influence flow infinitely. According to Theorem 5.5.7, for the regular community, all row vectors of the limiting matrix $\mathbf{T}^\infty$ are the same. This means that on every agent in the regular community, an agent has the same influence in the limit. Given an agent j in the regular community G, for all $i, k \in G$, $\mathbf{T}^\infty_{ij} = \mathbf{T}^\infty_{kj}$. Therefore, the number in the constant column vector $\mathbf{T}^\infty_{*j}$ can be taken as the agent j's influence on the whole regular community. It represents how influential the agent j is in the regular community. Moreover, notice that the row vector in $\mathbf{T}^\infty$ is a strictly positive probability vector. Hence what Proposition 5.7.1 conveys is that the regular community tends to believe Q if and only if the members in the regular community who believe Q are more influential than those who do not. For convenience, we will call the limiting matrix of the regular community's trust matrix "*influence matrix*" and its row vector "*influence vector*". The influence vector will be denoted by $\mathbf{u}$.

Keeping the new way of understanding the regular community's potential belief in mind, we introduce the language for the logic.

5.7.2. Definition. Let G be a finite set of agents, At be a set of atomic propositions. The language $\mathcal{L}_{B\succcurlyeq}$ is given by the following grammar:

$$\varphi ::= p \mid C \succcurlyeq D \mid \neg\varphi \mid \varphi \wedge \varphi \mid B_i\varphi$$

where $p \in \mathsf{At}$ and $C, D \subseteq G$.

B_i is the belief operator for each single agent. $\succcurlyeq$ is a comparison operator, expressing that group C has more impact on the whole regular community than group D.

We do not explicitly include the regular community's potential belief as an operator in the language. But the language is expressive enough to define the potential belief operator as we will prove later in this section. Before that, we first give the truth conditions for the language in a Kripke-DeGroot model.

5.7.3. Definition. Given a Kripke-DeGroot model $\mathcal{M} = (\mathcal{KD}, V)$ where $\mathcal{KD}$ is a Kripke-DeGroot frame and $V : At \to 2^W$ is a valuation function, and a possible world w in W,

$\mathcal{M}, w \vDash p$	iff	$w \in V(p)$
$\mathcal{M}, w \vDash C \succcurlyeq D$	iff	$\sum_{i \in C} \mathbf{u}_i \geq \sum_{i \in D} \mathbf{u}_i$
$\mathcal{M}, w \vDash \neg\varphi$	iff	$\mathcal{M}, w \nvDash \varphi$
$\mathcal{M}, w \vDash \varphi \wedge \psi$	iff	$\mathcal{M}, w \vDash \varphi$ and $\mathcal{M}, w \vDash \psi$
$\mathcal{M}, w \vDash B_i\varphi$	iff	$R_i(w) \subseteq [\![\varphi]\!]$

The semantic truth of the formula $C \succcurlyeq D$ relies on the influence vector $\mathbf{u}$ rather than directly on the trust matrix. As we have analysed, $\succcurlyeq$ compares which group of agents is more influential. Then how can we make use of the operators $\succcurlyeq$ and B_i to define the operator for the regular community's potential belief?

Recall the definition of the vector $\mathbf{b}^Q$ in Definition 5.6.2 where Q is a subset of the set of possible worlds. A more general form of this kind of characteristic vector is as follows:

5.7.4. Definition. Given any group of agent G, any $i \in G$ and any subset C of G,

$$\mathbf{b}_i^C = \begin{cases} 1 & i \in C \\ 0 & \text{otherwise} \end{cases} \tag{5.14}$$

If we take $\{i \in G \mid f(i) \subseteq Q\}$ as the set C in the definition, we get Definition 5.6.2.

To define the regular community's potential belief in language $\mathcal{L}_{B\succcurlyeq}$, we should express the sufficient and necessary condition for it in Proposition 5.7.1. Although the condition in Proposition 5.7.1 is not expressible directly in language $\mathcal{L}_{B\succcurlyeq}$, a second thought indicates that we can reach the goal by taking the following detour.

5.7.5. Corollary. *Given a regular KDF $\mathcal{KD}$, the regular community G tends to believe Q if and only if there exists a subset $C \subseteq G$ such that $C \subseteq \{i \in G \mid f(i) \subseteq Q\}$ and*

$$\mathbf{T}_{i*}^{\infty}\mathbf{b}^C > 0.5$$

We express the condition in the above corollary in $\mathcal{L}_{B\succcurlyeq}$ by

$$B_G\varphi := \bigvee_{C\subseteq G}((C \succ \overline{C}) \wedge \bigwedge_{i\in C} B_i\varphi)$$

where $\overline{C} = G \setminus C$ and $C \succ \overline{C} := (C \succcurlyeq \overline{C}) \wedge \neg(\overline{C} \succcurlyeq C)$.

5.7.6. Proposition. *Given any regular Kripke-DeGroot model $\mathcal{M}$ and any $w \in W$, $\mathcal{M}, w \vDash B_G\varphi$ if and only if there exists a subset $C \subseteq G$ such that $C \subseteq \{i \in G \mid f(i) \subseteq [\![\varphi]\!]\}$ and*

$$\mathbf{T}_{i*}^{\infty}\mathbf{b}^C > 0.5$$

Note that the regular community G is finite. So we can express the existence of a subset of G by taking the disjunction. The expression of potential group belief in the language $\mathcal{L}_{B\succcurlyeq}$ indicates that the logical properties of potential group belief hinge on the logical properties of the influence comparison operator $\succcurlyeq$ and the single agent's belief operator B_i. So the axiomatization of the logic with language $\mathcal{L}_{B\succcurlyeq}$ can help us understand potential group belief better.

For the axiomatization, notice that $\succcurlyeq$ can never operate on any propositions. In fact, the formulas of the form $C \succcurlyeq D$ can be seen as atomic propositions with their inner structures specified in the language. Since it is clear that the axiom system for the operator B_i is the K system, the only task is to figure out what axioms are enough for characterising $\succcurlyeq$.

For $\succcurlyeq$, the task becomes easy once it is realized that the operator $\succcurlyeq$ is essentially the qualitative probability operator studied in Segerberg (1971) and Gärdenfors (1975). The only difference is that the probability in the Kripke-DeGroot model is distributed among the group of agents and is strictly positive, while in Segerberg (1971) and Gärdenfors (1975), it is distributed among possible worlds and is not necessarily strictly positive.

Therefore, we propose the axiom system KS in Table 5.1 for the logic of regular community's potential belief.

The axiom schema **Scott** in the axiom system KS says that given two sequence of subgroups of G: $C_0, \dots, C_n$ and $D_0, \dots, D_n$, if for all $i \in G$, the number of subgroups i belongs to in the C sequence is the same as that in the D sequence[7], then the formula $\bigwedge_{i=0}^{n-1}(C_i \succcurlyeq D_i) \to (D_n \succcurlyeq C_n)$ is derivable. It is shown to be necessary and together with the other two axioms **SP** and **CO** sufficient for constructing a probability measure on a finite sample space (Kraft, Pratt and Seidenberg, 1959; Scott, 1964) satisfying certain constraints imposed by the order $\succcurlyeq$.

Propositional tautologies and Modus Ponens
K: $B_i\varphi \to (B_i(\varphi \to \psi) \to B_i\psi)$
Necessitation rule: if φ is derivable, then $B_i\varphi$ is derivable
$\succcurlyeq$**4**: $C \succcurlyeq D \to B_i(C \succcurlyeq D)$ $\succcurlyeq$**5**: $\neg(C \succcurlyeq D) \to B_i\neg(C \succcurlyeq D)$
SP: $C \succ \varnothing$ for $C \neq \varnothing$
CO: $(C \succcurlyeq D) \vee (D \succcurlyeq C)$
Scott: if $\lvert\{k \mid i \in C_k \text{ and } 0 \leq k \leq n\}\rvert = \lvert\{l \mid i \in D_l \text{ and } 0 \leq l \leq n\}\rvert$ for all $i \in G$, then $\bigwedge_{i=0}^{n-1}(C_i \succcurlyeq D_i) \to (D_n \succcurlyeq C_n)$ is derivable

Table 5.1: Axiom system KS for the logic of the regular community's potential belief

5.7.7. Theorem. *The axiom system KS is strongly complete and sound with respect to the class of the regular Kripke-DeGroot models.*

Proof:
The proof can be found in Appendix D.5. □

5.7.8. Remark. The logic for the single agent's belief in KS is K rather than KD45 as we advocated in the previous chapters. This is not an essential change.

[7] Note that this is a meta-level fact about the language which is not dependent on the semantics. So whether it holds can be checked by counting the times of appearance of i in each of the two sequences of subgroups and then comparing the two numbers

The only reason for doing so is to stay as general and basic as possible. We can make B_i a KD45 operator by imposing corresponding conditions on the accessibility relation R_i, which will not pose any difficulties of retaining and proving the completeness and soundness of the logic.

The main confusion may come from the axioms $\succcurlyeq$**4** and $\succcurlyeq$**5**. They seem to be too strong and are not necessary for keeping the update iterating. In contrast, what seems to be necessary for the update is not required by our setting, that is, each group member's correct belief of her/his neighbours' beliefs.

We agree that these epistemic aspects can be pivotal to the diffusion process and thus affect the notion of potential group belief. We do not try to argue philosophically for our modelling choices when it concerns these epistemic aspects. The only reason for making these choices is to simplify the setting and thus explicate the idea of introducing the Kripke semantics.

In the next section, we will pick up the threads of the discussion about the conjunction rule. The failure of the conjunction rule has been a recurrent theme in this thesis. Once again, it happens to potential group belief. How should we save it this time? We will see that the cure we propose reveals a connection with the theory of judgement aggregation.

5.8 Discussion

5.8.1 Saving the failure of the conjunction rule

Let us first present a counterexample to illustrate why the closure under conjunction fails for potential group belief.

5.8.1. Example. Given a Kripke-DeGroot model $(\{pq, p\overline{q}, \overline{p}q, \overline{pq}\}, \mathbf{T}, f, V)$, where

- $\mathbf{T} = \begin{bmatrix} 0.3 & 0.3 & 0.4 \\ 0.3 & 0.3 & 0.4 \\ 0.3 & 0.3 & 0.4 \end{bmatrix}$
- $f(1) = \{pq\}, f(2) = \{pq, p\overline{q}\}, f(3) = \{pq, \overline{p}q\}$
- $V(p) = \{pq, p\overline{q}\}, V(q) = \{pq, \overline{p}q\}$

Because

$$\mathbf{T}^\infty = \mathbf{T} = \begin{bmatrix} 0.3 & 0.3 & 0.4 \\ 0.3 & 0.3 & 0.4 \\ 0.3 & 0.3 & 0.4 \end{bmatrix}$$

and

$$\mathbf{b}^{[\![p]\!]} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad , \quad \mathbf{b}^{[\![q]\!]} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \quad , \quad \mathbf{b}^{[\![p \wedge q]\!]} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad .$$

According to the definition of the regular community's potential belief, $B_G p$ and $B_G q$ hold in the model while $B_G(p \wedge q)$ does not.

The reason behind the failure is clear. Taking the conjunction of two propositions means that fewer agents would believe it than those who believe either of the conjuncts. And fewer agents means a less influential group, which is not enough for the conjunction to be the group's potential belief.

Since Section 5.6, we have restricted our attention to the regular community. This enables us to take the group's potential beliefs as the propositions which are believed by everyone in an influential enough group. Recall that it is decided by the influence vector (the row vector in $\mathbf{T}^{\infty}$) how influential a subgroup of agents is. The threshold for being influential enough is 0.5 . Under this interpretation of "influential enough", potential group belief fails to satisfy the conjunction rule. Is there any other interpretation under which the new notion of group belief follows the conjunction rule? The following definition fulfills the requirement.

5.8.2. DEFINITION. Given a regular Kripke-DeGroot frame, the regular community has *stable belief* of $Q \subseteq W$ if and only if there is a non-empty subgroup of agents G_{sub} such that for all $i \in G_{\text{sub}}$,

$$\mathbf{u}_i > \sum_{j \notin G_{\text{sub}}} \mathbf{u}_j \text{ and } f(i) \subseteq Q$$

where $\mathbf{u}$ is the influence vector.

The definition of the *stable group belief* redefines the term "influential enough". For potential group belief, "influential enough" simply means the whole subgroup's influence is larger than 0.5, while for stable group belief, "influential enough" means each member in the subgroup is more influential than the outsiders altogether. The requirement for "influential enough" becomes much stronger. Thus the group's stable belief implies the group's potential belief. Moreover, the set of group's stable beliefs is closed under under conjunction and is strongly consistent, because there must be a smallest "influential enough" set of agents in the regular community according to Definition 5.8.2, which completely decides the group's stable belief.

5.8.3. PROPOSITION. *Given a regular Kripke-DeGroot frame $\mathcal{KD} = (W, \mathbf{T}, f)$,*

1. *for all $Q \subseteq W$, if the given group G has stable belief of Q, then it has potential belief of Q.*

2. *for all $Q, S \subseteq W$, if the given group G has stable belief of Q and has stable belief of S, then it has stable belief of $Q \wedge S$.*

Depending on the influence vector of the regular community which decides how large is the "influential enough" set of agents, the group's stable beliefs can be dictatorial or extremely democratic. If there is a super influential member in the regular community such that (s)he is more influential than the others altogether, then the group's stable beliefs result from the dictatorship. If all members are equally influential, then the group's stable beliefs have to be achieved by a consensus of all the members. The most critical point here is that the influence vector is decided by the current structure of the social network. This structure of social network, nevertheless, is decided by each member in the group, that is, how each member in the group distributes her trust among the group members. No matter whether the final "influential enough" set of agents includes only one agent or every agent, the way of picking out this set of agents respects each group member's choices equally.

At last, we note that the regular community's stable belief is expressible in the language $\mathcal{L}_{B\succcurlyeq}$ with respect to the class of regular Kripke-DeGroot models.

5.8.4. Proposition. *Given a regular Kripke-DeGroot model $\mathcal{M}$, for all $\varphi \in \mathcal{L}_{B\succcurlyeq}$, the group has stable belief of φ in a possible world w if and only if*

$$\mathcal{M}, w \vDash \bigvee_{C \subseteq G, C \neq \varnothing} \bigwedge_{i \in C} ((\{i\} \succ \overline{C}) \wedge B_i\varphi)$$

5.8.2 Connection with judgement aggregation

The terms "democratic" and "dictatorial" used in the discussion of stable group belief have indicated a possible connection with the theory of judgement aggregation (List and Pettit, 2002; Dietrich and List, 2007; Dokow and Holzman, 2010; List, 2012). Indeed, the connection can already be spotted when we start viewing the notion of potential group belief through the influence matrix. The group's potential belief of Q depends on whether the group members who believe Q are influential enough. It can be seen as a generalisation of the majority rule used in voting. The connection becomes even more obvious when the failure of the conjunction rule for potential group belief was brought up.

The closure failure for the regular community's potential belief is quite similar to the *discursive dilemma* (Pettit, 2001) or the *majority inconsistent* (Dietrich and List, 2014) in judgment aggregation. The same observation is also made in Dietrich and List (2014) where they do a "comprehensive study of the lessons that we can learn for belief binarisation from the large terrain of aggregation-theoretic impossibility and possibility results" (p.3).

In the context of a single agent, there are several proposals as to how the task of belief binarisation from the agent's subjective probability can be done. For example, Leitgeb's *stability theory of belief* (Leitgeb, 2014) and Lin and Kelly's *camera shutter rule* (Lin and Kelly, 2012*a*,*b*). It has been suggested in Dietrich and List (2014) that these proposals about connecting single agent's probabilistic information and qualitative belief can carry over to judgment aggregation theory. And some papers explore further in this direction and flesh out details about how the carry-over can work. For example, in Cariani (2016), the author builds up the structural analogy between Leitgeb's *P-stability-based rule* in his stability theory and a judgment aggregation rule, the *closed local supermajorities* rule ($\mathbf{LS}^+$).

The notion of stable group belief proposed in Definition 5.8.2 is closely related to Leitgeb's stability theory. If we see each agent in the regular community as a possible world and take the influence vector $\mathbf{u}$ as the probability distribution, then those subsets of G which satisfy the condition $\mathbf{u}_i > \sum_{j \notin G_{\text{sub}}} \mathbf{u}_j$ are the so-called *P-stable* propositions (Leitgeb, 2014, Definition 1).

We have shown that stable group belief can be dictatorial or democratic, depending on how the influence is distributed among group members. A new message conveyed in stable group belief (also in potential group belief) which social choice theorist may find interesting is that the influence vector can be seen as the result of aggregating each group member's trust assignment. This aggregation is inherent in the social network of the group. So even if the final step of deciding the group's belief can be dictatorial, it is the choice of the group, inherent in its network structure.

5.9 Conclusion

In this Chapter, we propose to take the group's tendency towards consensus as its belief, where the tendency is decided by the interaction of group members through their social network. As the DeGroot model, we represent the social network by a trust matrix and use weighted average to capture the effect of the interaction through the social network on each agent's belief. Different from the DeGroot model, we restrict our attention to each agent's binary belief rather than degrees of belief. Moreover, in the IBDM, the effect of the interaction does not decide how the agent changes her degree of belief but only indicates the probability of the agent changing her binary belief. This new perspective leads to the qualitative notion of group belief based on the probabilistic measure of the group members' belief update under social influence.

The Markov chain theory helps characterise the process running behind the notion of potential group belief. In turn, the shift of perspective and the analysis of potential group belief also shed light on our understanding of the Markov chain theory.

We discussed the failure of the conjunction rule again and proposed a new notion of group belief,in order to avoid this failure. The tension between believing more and believing more consistently reappears. Recall that in Chapter 2, the notion of justified belief ensures that the agent believes more consistently while the notion of grounded belief allows the agent to have more beliefs. For group belief, the notion of stable group belief ensures more consistency while the notion of potential belief allows for more beliefs. Should the group choose to ensure more consistency or accept more information? We tend to think that the answer depends on practical considerations, which go beyond the scope of this thesis.

After illustrating the different sides of the notion of potential group belief, at the end of this Chapter, it may be helpful to see the opposite side of the notion.

First, the notion of potential group belief is *not* based on a probability measure on *the space of possible worlds*. For a single agent, it is usually stipulated that her subjective probability is distributed among a set of possible worlds. It implies that if the proposition Q has probability x then its negation has probability $1-x$. There is no such probability measure for the group distributed among the possible worlds. $\mathbb{T}^{\infty}_{\mathbf{b}^{Q}\mathbf{1}} = x$ does not imply that $\mathbb{T}^{\infty}_{\mathbf{b}^{\overline{Q}}\mathbf{1}} = 1 - x$. The probability, which is the probability in a Markov process, is distributed among $\mathfrak{b}$ (the set of all possible group belief states). For each proposition, the probability distribution is different.

Second, the group's potential belief is not the group's destiny; it is just a tendency, and the tendency changes whenever the current situation changes.

Chapter 6

Conclusion

As a logical study of belief, this dissertation echoes the recent trend of looking deep into the base of belief (van Benthem and Pacuit, 2011; van Benthem, Fernández-Duque and Pacuit, 2012; van Benthem, Fernández-Duque and Pacuit, 2014; Baltag, Renne and Smets, 2014; Baltag, Fiutek and Smets, 2016; Baltag, Bezhanishvili, Özgün and Smets, 2016*a*; Baltag and Occhipinti, 2017) and the trend of putting more emphasis on its social aspects (Seligman, Liu and Girard, 2013; Liu, Seligman and Girard, 2014; Liu and Lorini, 2016; Baltag, Christoff, Rendsvig and Smets, 2016; Christoff, 2016). The focus of this dissertation, nevertheless, is more directed towards filling the gap between evidence and belief, the gap between single-agent belief and group belief and on what bridges these gaps – reasoning, no matter whether it is a single agent's reasoning or a social design which reflects the group's network structure.

The whole dissertation characterises several notions of belief and their relations to their bases – evidence/arguments or group members' beliefs – in different sound and complete logical systems. These logics bring together several different fields, for example, the topological semantics for evidence and formal argumentation theory in the logic of argument and belief; Markov chain theory and the Kripke semantics in the logic of the regular community's potential belief. The integration of these techniques enables us to penetrate into the doxastic earth and reveal not only different doxastic attitudes but also different types of reasoning running between and behind them.

In Chapter 2, we started with a review of the notion "justified belief" in Baltag, Bezhanishvili, Özgün and Smets (2016*a*), which is a topological version of "evidence-based belief" in van Benthem and Pacuit (2011). Then we proposed the notion "grounded belief", which aims to lower the high standard of justified belief. Argumentational reasoning conceptually fulfils this task. Technically, the representation of argumentational reasoning is made possible by integrating formal argumentation theory (Dung, 1995) into the topological semantics for evidence. The sound and complete logic of argument and belief developed in Section 2.5

characterises grounded belief and its relations with justified belief and arguments.

The application of the logic of argument and belief can also be found in Chapter 3, which is used to reason about the relations between three notions of belief – justified belief, grounded belief and full-support belief. Full-support belief brings back full consistency which grounded belief lacks while keeping the standard of belief not as high as justified belief by appealing to default reasoning. Full-support belief, therefore, serves as an answer to the question raised at the beginning of this dissertation – whether a balance can be struck between believing more and believing more consistently. Such a possibility of a balance also justifies the full consistency of a group member's belief in Chapter 4 and 5.

In Chapter 4, we turned to study group belief. Argumentation becomes a way of resolving conflicts between group members rather than between different arguments within a single agent's mind. The logic of group belief and argumentation explicitly expresses attack relations with respect to different issues, which add the second dimension to the logic. This second dimension plays a vital role in shaping a group's belief, as embodied in the notion of argumentation-based group belief.

In Chapter 5, instead of an adversarial relationship between group members, a more friendly relation is on focus – trust. The indeterministic binary DeGroot model bridges the gap between group members' qualitative beliefs and a stochastic process depending on the group's trust matrix. The Kripke-DeGroot model witnesses the meeting of the Kripke semantics and the DeGroot model and makes a logical study of potential group belief possible.

Compared to the concept of everyone's belief and distributed belief which we introduced in Section 1.2, both argumentation-based group belief and potential group belief are less extreme. They are both easier to achieve than everyone's belief and not as inconsistent as distributed belief. However, for group belief, the problem is not only about consistency but also about the morality of the mechanism of achieving it. If we take it for granted that under certain conditions, democracy is always more desirable than dictatorship, then it seems that "majority inconsistency" (Dietrich and List, 2014), which drives an impossibility theorem for judgement aggregation, is unavoidable. Although we see that stable group belief in Chapter 5 is fully consistent while not as stringent as everyone's belief, it allows dictatorship. There is no way of dealing with group belief purely from a logician's perspective. That is why we tend to leave the issue open and let practical need decide the right notion of group belief.

Despite the similarity of their logical properties, there is an apparent discrepancy between the two ways of defining group belief. Argumentation-based group belief relies on an explicit deliberative process, while potential group belief depends on implicit and potential opinion diffusion. The relationship between group members in argumentation-based group belief is more adversarial. The relationship between group members in potential group belief is less discursive or deliberative. Is there any possibility of reconciling these two ways of defining group belief?

Besides the discrepancy between the two notions of group belief, there is a divergence between the views on how the single agent forms and changes her belief, which are embodied in the notions of single-agent belief (grounded belief and full-support belief) and potential group belief respectively. While grounded belief and full-support belief emphasise the role of the agent's reasoning and evidence, potential group belief seems to underline the role of social impact. In Johan van Benthem's term (van Benthem, 2015*a*), this is a divergence between "high rationality" and "low rationality". The divergence between these two views does not mean that they are irreconcilable. The problem is how they can be integrated into one framework and coherently interact with each other.

The work in this dissertation pays more attention to how belief is formed based on available information, no matter whether it is in a highly rational way or not. In this way it complements dynamic epistemic logic which mainly studies how belief and knowledge change when new information is received (van Benthem, 2007). It is worthwhile to pursue further how these two different perspectives can complement each other and thus bring each other new insights. After all, reasoning to believe is not a static scene but a dynamic process in essence.

At last but not least, after getting a closer look at the different notions of belief, it may be helpful to reflect on the relationship between belief and other propositional attitudes such as preference (van Benthem and Liu, 2007; Liu, 2008) and knowledge (Holliday, 2012; Egré, 2017). It will be beneficial to understand further how our perspective and results on belief in this dissertation can be applied to related problems in decision theory and (social) epistemology.

Appendix A

Appendix of Chapter Two

A.1 Proof of Theorem 2.4.6

Soundness is straightforward based on what we have proved in Section 2.5; for completeness, the proof uses a modal equivalence result.

First, define a *belief neighborhood model* M as a uniform neighborhood model (W, N_{B}, V) where the neighborhood function $N_{\mathsf{B}} \subseteq 2^W$ satisfies the following conditions:

- $W \in N_{\mathsf{B}}$ (N_{B} contains the unit);
- if $b \in N_{\mathsf{B}}$, then $b' \in N_{\mathsf{B}}$ for all b' such that $b \subseteq b'$ (N_{B} is closed under supersets);
- if $b \in N_{\mathsf{B}}$, then $W \setminus b \notin N_{\mathsf{B}}$ (N_{B} does not contain the complement of any of its elements).

In such structures, the semantic interpretation of a modality for this *neighborhood-based belief* is given by

$$\mathsf{M}, w \vDash \mathsf{B}\varphi \quad \text{iff} \quad [\![\varphi]\!]_{\mathsf{M}} \in N_{\mathsf{B}}$$

Given a topological argumentation model, it is easy to build a point-wise modally equivalent belief neighborhood model: it is enough to define the neighborhood function N_{B} as LFP_τ plus all its elements' supersets ($N_{\mathsf{B}} := \{b \in 2^W \mid \mathsf{f} \subseteq b$ for some $\mathsf{f} \in \mathsf{LFP}_\tau\}$). The readers can check that the model as constructed above is indeed a belief neighborhood model and it is point-wise modally equivalent to the given topological argumentation model.

We still need to prove the other direction of modal equivalence, so that we can make use of the completeness result of the axiom system EMND45 with respect to belief neighbourhood models (Chellas, 1980, Chapter 8 and 9).

A.1.1. LEMMA. *For every belief neighborhood model* $\mathsf{M} = (W, N_{\mathsf{B}}, V)$ *there is a topological argumentation model* $\mathcal{M} = (W, E_0, \tau, \twoheadleftarrow, V)$ *with the same domain and atomic valuation such that* $\mathcal{M}$ *and* M *are point-wise modally equivalent with respect to the language* $\mathcal{L}$*, that is, for all* $\varphi \in \mathcal{L}$ *and all* $w \in W$*,*

$$\mathcal{M}, w \vDash \varphi \textit{ if and only if } \mathsf{M}, w \vDash \varphi \ .$$

Proof:
Given an arbitrary belief neighbourhood model $\mathsf{M} = (W, N_{\mathsf{B}}, V)$, let

$$\mathcal{M}_{\mathsf{M}} = (W, E_0, \tau, \twoheadleftarrow_{N_{\mathsf{B}}}, V)$$

be the topological argumentation model that shares domain and atomic valuation with M, and in which the family of pieces of evidence is given by the singletons in W ($E_0 := \{\{w\} \in 2^W \mid w \in W\}$) and thus the generated topology is the power set of the domain ($\tau = 2^W$). Moreover, define the attack relation $\twoheadleftarrow_{N_{\mathsf{B}}}$ as

$$t \twoheadleftarrow_{N_{\mathsf{B}}} t' \quad \text{iff} \quad \begin{cases} t \cap t' = \varnothing \text{ and } t \notin N_{\mathsf{B}} & \text{when } t' \neq \varnothing \\ t = \varnothing & \text{when } t' = \varnothing \end{cases}$$

for every $t, t' \in \tau$, so a non-empty t' attacks a non-empty t if and only if they are in conflict and t is not in N_{B}, and while the empty set does not attack non-empty sets, it is attacked by everybody (including itself). We first verify that this model is a topological argumentation model.

We only need to show that $\twoheadleftarrow_{N_{\mathsf{B}}}$ satisfies the three requirements in Definition 2.3.1.

- Take any $t_1, t_2 \in \tau$.

  ($\Rightarrow$) Suppose $t_1 \twoheadleftarrow_{N_{\mathsf{B}}} t_2$, by $\twoheadleftarrow_{N_{\mathsf{B}}}$'s definition, $t_1 \cap t_2 = \varnothing$.

  ($\Leftarrow$) Suppose $t_1 \cap t_2 = \varnothing$. If $t_2 = \varnothing$, $t_2 \twoheadleftarrow_{N_B} t_1$

  If $t_1 \in N_{\mathsf{B}}$; now, from $t_1 \cap t_2 = \varnothing$ it follows that $t_1 \subseteq W \setminus t_2$. By closure under superset and consistency of N_{B}, $t_2 \notin N_{\mathsf{B}}$. So by definition, $t_2 \twoheadleftarrow_{N_{\mathsf{B}}} t_1$.

  If $t_1 \notin N_{\mathsf{B}}$, by $\twoheadleftarrow_{N_{\mathsf{B}}}$'s definition, it follows that $t_1 \twoheadleftarrow_{N_{\mathsf{B}}} t_2$ when $t_2 \neq \varnothing$ or $t_2 \twoheadleftarrow_{N_{\mathsf{B}}} t_1$ when $t_2 = \varnothing$.

- Take $t, t_1, t_1' \in \tau$ such that $t_1 \twoheadleftarrow_{N_{\mathsf{B}}} t$ and $t_1' \subseteq t_1$. $t_1 \twoheadleftarrow_{N_{\mathsf{B}}} t$ implies $t_1 \notin N_{\mathsf{B}}$. So $t_1' \notin N_{\mathsf{B}}$ should be the case, otherwise by $t_1' \subseteq t_1$ and N_{B}'s closure under supersets we would get the contradictory $t_1 \in N_{\mathsf{B}}$. But $t_1 \twoheadleftarrow_{N_{\mathsf{B}}} t$ implies $t \cap t_1 = \varnothing$, which together with $t_1' \subseteq t_1$ yields $t \cap t_1' = \varnothing$. Hence, from $\twoheadleftarrow_{N_{\mathsf{B}}}$'s definition, $t_1' \twoheadleftarrow_{N_{\mathsf{B}}} t$.

- The attack conditions on $\varnothing$ are 'embedded' in $\twoheadleftarrow_{N_{\mathsf{B}}}$'s definition.

Second, we show that this topological argumentation model is point-wise modally equivalent to the given belief neighbourhood model. For this purpose, we only need prove the following fact.

A.1.2. LEMMA.

$$\mathsf{LFP}_{\tau} = N_{\mathsf{B}} \ .$$

Proof:
Since for all $t \in N_{\mathsf{B}}$, t is not attacked by any other arguments, it follows that $N_{\mathsf{B}} \subseteq d(N_{\mathsf{B}})$ and N_{B} is a subset of LFP_{τ}. So, to prove that N_{B} is the least fixed point, we only need to prove that N_{B} is a fixed point, which is equivalent in this case to $d(N_{\mathsf{B}}) \subseteq N_{\mathsf{B}}$.

Assume that $t \notin N_{\mathsf{B}}$. Then it follows that $t \twoheadleftarrow_{N_{\mathsf{B}}} W \setminus t$ by definition of $\twoheadleftarrow_{N_{\mathsf{B}}}$ and there is no $t' \subseteq t$ such that $t' \in N_{\mathsf{B}}$ by the closure under superset of N_{B}.

So for all $t'' \in N_{\mathsf{B}}$, $t'' \cap W \setminus t \neq \varnothing$ and thus t'' does not attack $W \setminus t$. Hence $t \notin d(N_{\mathsf{B}})$. □

□

We have built up the pointwise modal equivalence between the topological argumentation model and the belief neighbourhood model. The remaining work is to prove the completeness of the system EMND45 with respect to the belief neighbourhood model. The details can be found in (Chellas, 1980, Chapter 8 and 9).

A.2 Proof of Theorem 2.5.7

The proof for the soundness is straightforward, and the nontrivial part of the work has been done. So we focus on the proof of the strong completeness, which is equivalent to the satisfiability of an arbitrary consistent set of formulas in $\mathcal{L}_{\forall \mathcal{T} \Box}$ in a *topological argumentation model*.

All twists of the proof are caused by the operator $\mathcal{T}$ in the language and the attack relation in the constructed model, which are mainly dealt with in the second stage of this proof.

A.2.1 First stage

Given a set of formulas in $\mathcal{L}_{\forall \mathcal{T} \Box}$, denoted by Φ_0, we first construct a *quasi-a-model* where Φ_0 is satisfiable.

Let MCS be the family of all maximally ABS-consistent sets. By a slightly modified version of the Lindenbaum Lemma (whose proof is standard), which says that every ABS-consistent set of formulas in $\mathcal{L}_{\forall \mathcal{T} \Box}$ can be extended to a maximally consistent one, it can be shown that Φ_0 can be extended to a maximally ABS-consistent set, denoted by Φ.

For each set of formulas $\Gamma \subseteq \mathcal{L}_{\forall \mathcal{T} \Box}$, let $\Gamma^{\bigcirc} = \{\phi \in \mathcal{L}_{\forall \mathcal{T} \Box} \mid \bigcirc\phi \in \Gamma\}$ where $\bigcirc$ can be either $\forall$ or else $\Box$ or else $\mathcal{T}$.

A.2.1. Definition (Canonical quasi-a-model). The canonical quasi-a-model is $\mathcal{M}^{\Phi} = (W^{\Phi}, \leqslant^{\Phi}, E_0^{\Phi}, \leftarrowtail^{\Phi}, V^{\Phi})$ where

- $W^{\Phi} = \{\Gamma \in \mathrm{MCS} \mid \Gamma^{\forall} = \Phi^{\forall}\}$;[1]
- for all $\Gamma, \Delta \in W^{\Phi}$, $\Delta \leqslant^{\Phi} \Gamma$ if and only if for all $\phi \in \mathcal{L}_{\forall \mathcal{T} \Box}$, $\Box\phi \in \Delta$ implies $\phi \in \Gamma$;
- $E_0^{\Phi} = \{\bigcup_{\Gamma \in S} \geqslant_{\Gamma}^{\Phi} \mid S \subseteq W^{\Phi}\} \setminus \{\varnothing\}$, where $\geqslant_{\Gamma}^{\Phi} = \{\Omega \in W^{\Phi} \mid \Gamma \leqslant^{\Phi} \Omega\}$;
- $V^{\Phi}(p) = \{\Gamma \in W^{\Phi} \mid p \in \Gamma\}$

The attack relation $\leftarrowtail^{\Phi}$ deserves special attention:

Let τ^{Φ} be the topology generated by E_0^{Φ}, for all $t, t' \in \tau^{\Phi}$, $t \leftarrowtail^{\Phi} t'$ if and only if

when $t' = \varnothing$**:** $t = \varnothing$*;*

when $t' \neq \varnothing$**:** $t \cap t' = \varnothing$ *and there is no* $\phi \in \mathcal{L}_{\forall \mathcal{T} \Box}$ *such that both* $\|\mathcal{T}\phi\| \subseteq t$ *and* $\exists \mathcal{T}\phi \in \Phi$*,*

where $\|\varphi\|^{\Phi} = \{\Gamma \in W^{\Phi} \mid \varphi \in \Gamma\}$

We are going to omit the superscript Φ for the notation introduced in the above definition in the subsequent proof when no confusion arises.

The next proposition includes the existence lemmas for $\Box$ and $\forall$, whose proof is standard and can be found in Özgün (2017, Proposition 5.6.19 and 5.6.20).

A.2.2. Proposition. *For all $\varphi \in \mathcal{L}$,*

1. $\exists\varphi \in \Phi$ *if and only if there exists* $\Delta \in W^{\Phi}$ *such that* $\varphi \in \Delta$
2. *for all* $\Delta \in W^{\Phi}$, $\Diamond\varphi \in \Delta$ *if and only if there exists* $\Gamma \in W^{\Phi}$ *such that* $\Delta \leqslant^{\Phi} \Gamma$ *and* $\varphi \in \Gamma$.

[1] This way of defining W is equivalent to the way adopted in Özgün (2017, Definition 5.6.17):

$$W^{\Phi} = \{\Gamma \in \mathrm{MCS} \mid \text{ for all } \phi \in \mathcal{L}_{\forall \mathcal{T} \Box}, \forall\phi \in \Phi \text{ implies } \phi \in \Gamma\} \ .$$

A.2.2 Second stage

The remaining part of the proof is directly related to the attack relation and the operator $\mathcal{T}$, which is the key part of this proof.

Before delving into the details, we review the definition of the attack relation $\twoheadleftarrow^{\Phi}$ and explain the idea behind the definition.

The idea behind the definition of $\twoheadleftarrow^{\Phi}$ is to make sure that

1. if $\mathcal{B}\phi \in \Phi$ and $\|\phi\| \subseteq t \in \tau^{\Phi}$ for some $\phi \in \mathcal{L}_{\forall\mathcal{T}\Box}$, then the argument t is not attacked; and

2. if an argument t in τ^{Φ} does not support any $\phi \in \mathcal{L}_{\forall\mathcal{T}\Box}$ such that $\mathcal{B}\phi \in \Phi$, then t is attacked by some other arguments.

1 and 2 are critical for us to pinpoint the least fixed point LFP in $\mathcal{M}^{\Phi}$ through the syntactical information given by Φ. While it is straightforward to see that 1 follows from the definition, it is unclear whether 2 also follows. This causes all the twists in Definition A.2.8 and Lemma A.2.11. Another uncertainty before we prove it is whether $\twoheadleftarrow^{\Phi}$ satisfies all three conditions we require for the attack relation.

Hence we first check that the constructed model in Definition A.2.1 is indeed a quasi-a-model. It is obvious that $\varnothing \notin E_0$ and $W \in E_0$. So we check that $\twoheadleftarrow^{\Phi}$ in the model satisfies the required conditions (Definition 2.3.5).

A.2.3. Lemma. *For the model in Definition A.2.1,*

- *for all $t, t' \in \tau$, $t \cap t' = \varnothing$ if and only if $t \twoheadleftarrow t'$ or $t' \twoheadleftarrow t$;*
- *for all $t_1, t_2, t_3 \in \tau$, if $t_1 \twoheadleftarrow t_3$ and $t_2 \subseteq t_1$, then $t_2 \twoheadleftarrow t_3$*
- *for all $t \in \tau \setminus \{\varnothing\}$, $\varnothing \twoheadleftarrow t$ and $t \not\twoheadleftarrow \varnothing$.*

Proof:
The first condition: We prove the direction from left to right. The proof of the other direction is trivial.

Assume that $t \cap t' = \varnothing$. Moreover, suppose that $t \not\twoheadleftarrow t'$ and $t' \not\twoheadleftarrow t$, then there is $\varphi \in \mathcal{L}_{\forall\mathcal{T}\Box}$ such that $\|\mathcal{T}\varphi\| \subseteq t$ and $\exists\mathcal{T}\varphi \in \Phi$, and there is $\psi \in \mathcal{L}_{\forall\mathcal{T}\Box}$ such that $\|\mathcal{T}\psi\| \subseteq t'$ and $\exists\mathcal{T}\varphi \in \Phi$. It follows that $\|\mathcal{T}\varphi\| \cap \|\mathcal{T}\psi\| = \varnothing$.

To reach the conclusion, it is enough to prove the following lemma.

A.2.4. Lemma. *For all $\varphi, \psi \in \mathcal{L}_{\forall\mathcal{T}\Box}$, $\|\mathcal{T}\varphi\| \cap \|\mathcal{T}\psi\| = \varnothing$ and $\exists\mathcal{T}\varphi \in \Phi$ and $\exists\mathcal{T}\psi \in \Phi$ lead to contradiction.*

Proof:
By axiom $\mathcal{T}\mathcal{T}\varphi \to \mathcal{T}\varphi$ and $\|\mathcal{T}\varphi\| \cap \|\mathcal{T}\psi\| = \varnothing$, we have $\|\mathcal{T}\mathcal{T}\varphi\| \cap \|\mathcal{T}\mathcal{T}\psi\| = \varnothing$. Notice that $\|\mathcal{T}\varphi\| \neq \varnothing$ and $\|\mathcal{T}\psi\| \neq \varnothing$ follow from $\exists\mathcal{T}\varphi \in \Phi$ and $\exists\mathcal{T}\varphi \in \Phi$

by Proposition A.2.2, which, together with axiom $\mathcal{T}\varphi \to \mathcal{T}\mathcal{T}\varphi$, implies that $\|\mathcal{T}\mathcal{T}\varphi\| \neq \varnothing$ and $\|\mathcal{T}\mathcal{T}\psi\| \neq \varnothing$. Using Lemma A.2.2 again, we can get $\exists\mathcal{T}\mathcal{T}\varphi \in \Phi$ and $\exists\mathcal{T}\mathcal{T}\psi \in \Phi$.

$\|\mathcal{T}\varphi\| \subseteq \|\neg\mathcal{T}\psi\|$ follows from $\|\mathcal{T}\varphi\| \cap \|\mathcal{T}\psi\| = \varnothing$, so $\forall(\mathcal{T}\varphi \to \neg\mathcal{T}\psi) \in \Phi$. Together with Fact 2.5.6 and S5 axioms and rules for $\forall$, it implies that $\forall(\mathcal{T}\mathcal{T}\varphi \to \mathcal{T}\neg\mathcal{T}\psi) \in \Phi$.

By using $\forall(\mathcal{T}\mathcal{T}\varphi \to \mathcal{T}\neg\mathcal{T}\psi) \in \Phi$ and $\exists\mathcal{T}\mathcal{T}\varphi \in \Phi$, we can derive $\exists\mathcal{T}\neg\mathcal{T}\psi \in \Phi$. By using axiom $\exists\mathcal{T}\varphi \to \neg\exists\mathcal{T}\neg\varphi$ and $\exists\mathcal{T}\mathcal{T}\psi \in \Phi$, we can derive $\neg\exists\mathcal{T}\neg\mathcal{T}\psi \in \Phi$. Contradiction. □

The second condition: Take any $t_1, t_2, t_3 \in \tau$ such that $t_1 \twoheadleftarrow t_3$ and $t_2 \subseteq t_1$. The case of $t_3 = \varnothing$ is trivial. Now consider $t_3 \neq \varnothing$. Then there is no $\varphi \in \mathcal{L}_{\forall\mathcal{T}\Box}$ such that $\|\mathcal{T}\varphi\| \subseteq t_1$ and $\exists\mathcal{T}\varphi \in \Phi$. Since $t_2 \subseteq t_1$, it follows that there is no $\varphi \in \mathcal{L}_{\forall\mathcal{T}\Box}$ such that $\|\mathcal{T}\varphi\| \subseteq t_2$ and $\exists\mathcal{T}\varphi \in \Phi$, i.e. $t_2 \twoheadleftarrow t_3$.

The proof of **the third condition** is trivial. □

Next, we collect a series of facts about the constructed quasi-a-model, which will be useful in our later proof.

A.2.5. FACT.

$$\tau = E_0 \cup \{\varnothing\}$$

A.2.6. FACT.

- If $\exists\Box\phi \in \Phi$, then $\|\Box\phi\| \in \tau$.
- If $\exists\mathcal{T}\phi \in \Phi$, then $\|\mathcal{T}\phi\| \in \tau$.

Proof:
The argument we use for proving two claims are similar. So we only provide the proof of the second claim.

Assume $\exists\mathcal{T}\phi \in \Phi$. Then by A.2.2, there is $\Gamma \in W$ such that $\mathcal{T}\phi \in \Gamma$. Take any $\Gamma \in W$ such that $\mathcal{T}\phi \in \Gamma$: by axiom $\mathcal{T}\phi \to \mathcal{T}\mathcal{T}\phi$ and $\mathcal{T}\phi \to \Box\phi$, then $\Box\mathcal{T}\phi \in \Gamma$. According to the definition of $\leqslant$, for all $\Delta \in W$ such that $\Gamma \leqslant \Delta$, $\mathcal{T}\phi \in \Delta$. So we have proved that for all $\Gamma \in W$ such that $\mathcal{T}\phi \in \Gamma$, $\geqslant_\Gamma \subseteq \|\mathcal{T}\phi\|$. Since $\|\mathcal{T}\phi\| \subseteq \bigcup_{\Gamma\in\|\mathcal{T}\phi\|} \geqslant_\Gamma$, it follows that $\|\mathcal{T}\phi\| = \bigcup_{\Gamma\in\|\mathcal{T}\phi\|} \geqslant_\Gamma$. So $\|\mathcal{T}\phi\|$ belongs to E_0 and thus τ by Fact A.2.5. □

A.2.7. FACT. For all $t \in \tau$ and all $\varphi \in \mathcal{L}_{\forall\mathcal{T}\Box}$, if $t \subseteq \|\varphi\|$, then $t \subseteq \|\Box\varphi\|$.

Proof:
The following fact is key to this proof: $t = \bigcup_{\Gamma\in t} \geqslant_\Gamma$. $t \subseteq \bigcup_{\Gamma\in t} \geqslant_\Gamma$ is easy to see. $t \supseteq \bigcup_{\Gamma\in t} \geqslant_\Gamma$ follows from the fact that if $\Gamma \in t$ then $\geqslant_\Gamma \subseteq t$.

Given the key fact that $t = \bigcup_{\Gamma \in t} \geqslant_\Gamma$ and the assumption that $t \subseteq \|\varphi\|$, it follows that for all $\Gamma \in t$, $\geqslant_\Gamma \subseteq \|\varphi\|$, which implies that $\Box\varphi \in \Gamma$ by Proposition A.2.2. Therefore, $t \subseteq \| \Box \varphi\|$. $\Box$

Starting from now, we try to locate the least fixed point LFP of the characteristic function d for the quasi-a-model by using the syntactic information, which will be helpful for the proof of the truth lemma.

A.2.8. Definition (Semi-acceptable and Acceptable). Let

$$C_1 = \{t \in \tau \mid \text{ there exists } \varphi \in \mathcal{L}_{\forall\mathcal{T}\Box} \text{ such that } \|\mathcal{T}\varphi\| \subseteq t \text{ and } \exists\mathcal{T}\varphi \in \Phi\} \ .$$

- An open $t \in \tau$ is semi-acceptable if and only if for all $\varphi \in \mathcal{L}_{\forall\mathcal{T}\Box}$ s.t. $t \subseteq \|\Box\varphi\|$, there exists $\psi \in \mathcal{L}_{\forall\mathcal{T}\Box}$ such that $\|\mathcal{T}\psi\| \subseteq \|\Box\varphi\|$ and $\exists\mathcal{T}\psi \in \Phi$.
- An open $t \in \tau$ is acceptable if and only if t is semi-acceptable and there is no $t' \in \tau$ s.t. for all $t'' \in C_1, t' \cap t'' \neq \varnothing$ and $t \cap t' = \varnothing$.

Denote the set of all acceptable opens in $\tau \smallsetminus C_1$ by C_2:

$$C_2 = \{t \in \tau \smallsetminus C_1 \mid t \text{ is acceptable}\} \ .$$

A quick observation is that all opens in C_1 are semi-acceptable. So they are also acceptable.

A.2.9. Fact. For all $t \in \tau$, if $t \in C_1$, then t is acceptable.

Moreover,

A.2.10. Lemma. *if $t \in \tau$ is semi-acceptable, then for all $t' \in C_1$, $t \cap t' \neq \varnothing$.*

Proof:
We prove that for all $\varphi \in \mathcal{L}_{\forall\mathcal{T}\Box}$ such that $\exists\mathcal{T}\varphi \in \Phi$, $t \cap \|\mathcal{T}\varphi\| \neq \varnothing$ given that t is a semi-acceptable open in τ. The lemma follows from this result.

Suppose that there exists $\varphi \in \mathcal{L}_{\forall\mathcal{T}\Box}$ such that $\exists\mathcal{T}\varphi \in \Phi$ and $t \cap \|\mathcal{T}\varphi\| = \varnothing$. Take any $\psi \in \mathcal{L}_{\forall\mathcal{T}\Box}$ such that $t \subseteq \|\Box\psi\|$. Then it follows that $t \subseteq \|\Box\psi \wedge \neg\mathcal{T}\varphi\|$ by $t \subseteq \|\Box\psi\|$ and $t \cap \|\mathcal{T}\varphi\| = \varnothing$. By Lemma A.2.7, $t \subseteq \|\Box(\Box\psi \wedge \neg\mathcal{T}\varphi)\|$. Together with the assumption that t is semi-acceptable, it implies that there is $\chi \in \mathcal{L}_{\forall\mathcal{T}\Box}$ such that $\|\mathcal{T}\chi\| \subseteq \|\Box(\Box\psi \wedge \neg\mathcal{T}\varphi)\|$ and $\exists\mathcal{T}\chi \in \Phi$. Since $\vdash \Box(\Box\psi \wedge \neg\mathcal{T}\varphi) \to \neg\mathcal{T}\varphi$, it follows that $\|\mathcal{T}\chi\| \subseteq \|\neg\mathcal{T}\varphi\|$ and thus $\|\mathcal{T}\chi\| \cap \|\mathcal{T}\varphi\| = \varnothing$. The rest of the proof can be done by using Claim A.2.4. $\Box$

We now prove the pivotal lemma in our proof.

A.2.11. Lemma. *Let $C = C_1 \cup C_2$,*

$$\mathsf{LFP} = C \ .$$

Proof:

We first prove the direction of $\supseteq$. Take any $t \in C$,

If $t \in C_1$, then there is $\varphi \in \mathcal{L}_{\forall\mathcal{T}\Box}$ such that $\|\mathcal{T}\varphi\| \subseteq t$ and $\exists\mathcal{T}\varphi \in \Phi$. By the definition of $\twoheadleftarrow$, there is no $t' \in \tau$ with $t \cap t' = \varnothing$ such that $t \twoheadleftarrow t'$. (Note that $t \in C_1$ implies that $t \neq \varnothing$, so $t \not\twoheadleftarrow \varnothing$.) Thus $C_1 \subseteq \mathsf{LFP}$.

If $t \in C_2$, take any $t' \in \tau$ such that $t \twoheadleftarrow t'$. According to the definition of $\twoheadleftarrow^{\Phi}$, $t' \cap t = \varnothing$. According to the definition of C_2 and $t \cap t' = \varnothing$, there is $t'' \in C_1$ such that $t' \cap t'' = \varnothing$. (Otherwise, t is not acceptable, because it violates the second condition of the definition of acceptability.) According to the definition of C_1, it follows that $t'' \not\twoheadleftarrow t'$. By condition 1 of $\twoheadleftarrow$ which is proved in Lemma A.2.3, it follows that $t' \twoheadleftarrow t''$. So for all $t \in C_2$, if $t \twoheadleftarrow t'$ then there is $t'' \in C_1$ such that $t' \twoheadleftarrow t''$. So $t \in d(C_1) \subseteq d(\mathsf{LFP}) = \mathsf{LFP}$.

Therefore, we have proved that $C \subseteq \mathsf{LFP}$

Next, we prove that $\mathsf{LFP}^{\Phi} \subseteq C$. Take an arbitrary $t \in \tau$ such that $t \notin C$, we try to prove that $t \notin \mathsf{LFP}$. If $t = \varnothing$, then $t \notin \mathsf{LFP}$, which is obvious, since $\varnothing \twoheadleftarrow \varnothing$. So we focus on the case where $t \neq \varnothing$.

$t \notin C$ implies that $t \notin C_1$, which in turn implies that there is no $\phi \in \mathcal{L}_{\forall\mathcal{T}\Box}$ such that $\|\mathcal{T}\phi\| \subseteq t$ and $\exists\mathcal{T}\phi \in \Phi$. It follows that for all $t' \in \tau$ such that $t \cap t' = \varnothing$, we have $t \twoheadleftarrow t'$ according to the definition of $\twoheadleftarrow$ and C_1.

We first prove that such a $t' \in \tau$ which satisfies $t \cap t' = \varnothing$ exists given $t \notin C$. Suppose not. Then for all $t' \in \tau$, $t \cap t' \neq \varnothing$.

Take an arbitrary $\varphi \in \mathcal{L}_{\forall\mathcal{T}\Box}$ such that $t \subseteq \|\varphi\|$. By Fact A.2.7, $t \subseteq \|\Box\varphi\|$. Take any $\Gamma \in W$. Since $\geqslant_{\Gamma} \in \tau$, it follows that $\geqslant_{\Gamma} \cap t \neq \varnothing$. By Proposition A.2.2 and the fact that there is $\Delta \geqslant \Gamma$ such that $\Delta \in t \subseteq \|\Box\varphi\|$, we have $\Diamond\Box\varphi \in \Gamma$.

Hence we have proved that for all $\Gamma \in W$ and all $\varphi \in \mathcal{L}_{\forall\mathcal{T}\Box}$ which satisfy $t \subseteq \|\varphi\|$, $\Diamond\Box\varphi \in \Gamma$. It follows that $\forall\Diamond\Box\varphi \in \Phi$ for all $\varphi \in \mathcal{L}_{\forall\mathcal{T}\Box}$ satisfying $t \subseteq \|\varphi\|$ by Proposition A.2.2. By axiom $\forall\Diamond\Box\varphi \to \exists\mathcal{T}\varphi \in \Phi$, $\exists\mathcal{T}\varphi \in \Phi$. By axiom $\mathcal{T}\varphi \to \Box\varphi$, $\|\mathcal{T}\varphi\| \subseteq \|\Box\varphi\|$.

Thus for all $\varphi \in \mathcal{L}_{\forall\mathcal{T}\Box}$ such that $t \subseteq \|\Box\varphi\|$, we have found a $\psi \in \mathcal{L}_{\forall\mathcal{T}\Box}$ (namely φ itself) such that $\|\mathcal{T}\psi\| \subseteq \|\Box\varphi\|$ and $\exists\mathcal{T}\psi \in \Phi$. So t is semi-acceptable. Moreover, since $t \cap t' \neq \varnothing$ for all $t' \in \tau$, there is no $t' \in \tau$ such that for all $t'' \in C_1$, $t \cap t'' \neq \varnothing$ and $t \cap t' = \varnothing$. We can now conclude that $t \in C_2 \subseteq C$. Contradiction.

Hence there must be $t' \in \tau$ such that $t \cap t' = \varnothing$. The rest of the proof is structured into two cases

- there is $t' \in \tau$ such that $t \cap t' = \varnothing$ and $t' \in C$;
- for all $t' \in \tau$ such that $t \cap t' = \varnothing$, $t' \notin C$

In the first case, take an arbitrary $t' \in \tau$ such that $t \cap t' = \varnothing$ and $t' \in C$. Then by the following three facts: (a) $t' \in C$, (b) $t \twoheadleftarrow t'$ as we have proved using the fact that $t \notin C$, and (c) $C \subseteq \mathsf{LFP}$ as we also have proved, it follows that $t \notin \mathsf{LFP}$, since LFP has to be conflict-free.

In the second case, it follows that for all $c \in C$, $t \cap c \neq \varnothing$. We need consider two sub-cases.

- t is semi-acceptable;
- t is not semi-acceptable.

If t is semi-acceptable, then there must be $t' \in \tau$ such that for all $c_1 \in C_1$, $t' \cap c_1 \neq \varnothing$ and $t \cap t' = \varnothing$. Otherwise, $t \in C_2$, which contradicts the assumption that $t \notin C$. Take any such t'. By $t \notin C$ and $t \cap t' = \varnothing$ and $t' \neq \varnothing$, it follows that $t \twoheadleftarrow t'$. Moreover, there is no $c_1 \in C_1$ such that $t' \twoheadleftarrow c_1$, because $t' \cap c_1 \neq \varnothing$; there is no $c_2 \in C_2$ such that $t' \twoheadleftarrow c_2$, because $t' \cap c_2 = \varnothing$, together with the fact that $t' \cap c_1 \neq \varnothing$ for all $c_1 \in C_1$, would imply that $c_2 \notin C_2$. Hence we have proved that if t is semi-acceptable, then there is $t' \in \tau$ such that $t \twoheadleftarrow t'$ and there is no $c \in C$ such that $t' \twoheadleftarrow c$, which means that $t \notin d(C)$.

Next we try to prove that $t \notin d(C)$ even if t is not semi-acceptable.

If t is not semi-acceptable, then there is $\varphi \in \mathcal{L}_{\forall \mathcal{T} \Box}$ such that $t \subseteq \| \Box \varphi \|$ and there is no $\psi \in \mathcal{L}_{\forall \mathcal{T} \Box}$ such that $\| \mathcal{T} \psi \| \subseteq \| \Box \varphi \|$ and $\exists \mathcal{T} \psi \in \Phi$. Take such a φ, denoted by φ_t. Then it follows that $\exists \mathcal{T} \varphi_t \notin \Phi$, because $\| \mathcal{T} \varphi_t \| \subseteq \| \Box \varphi_t \|$.

Now for each $c \in C_1$, take arbitrary $\varphi \in \mathcal{L}_{\forall \mathcal{T} \Box}$ such that $\| \mathcal{T} \varphi \| \subseteq c$ and $\exists \mathcal{T} \varphi \in \Phi$, denoted by φ_c. Next, we prove that $\| \Box \varphi_t \| \cap \| \mathcal{T} \varphi_c \| \neq \varnothing$ for every $c \in C_1$.

By axiom $\Box\varphi \to \Box \Box \varphi$ and theorem $\mathcal{T} \varphi \to \Box \mathcal{T} \varphi$ and K axiom for $\Box$, it follows that $\| \Box \varphi_t \wedge \mathcal{T} \varphi_c \| \subseteq \| \Box (\Box\varphi_t \wedge \mathcal{T} \varphi_c) \|$ for all $c \in C_1$, which by Proposition A.2.2 implies that $\forall((\Box\varphi_t \wedge \mathcal{T} \varphi_c) \to \Box(\Box\varphi_t \wedge \mathcal{T} \varphi_c)) \in \Phi$.

Up to now, we have proved that $\exists \mathcal{T} \varphi_t \notin \Phi$, and $\forall((\Box\varphi_t \wedge \mathcal{T} \varphi_c) \to \Box(\Box\varphi_t \wedge \mathcal{T} \varphi_c)) \in \Phi$.

By theorem $\mathcal{T} \varphi \leftrightarrow \mathcal{T} \Box\varphi$, $\exists \mathcal{T} \varphi_t \notin \Phi$ implies $\exists \mathcal{T} \Box\varphi_t \notin \Phi$. By applying axiom $\mathcal{T} \varphi \to \mathcal{T} \mathcal{T} \varphi$ to $\exists \mathcal{T} \varphi_c \in \Phi$, we have $\exists \mathcal{T} \mathcal{T} \varphi_c \in \Phi$. Together with $\forall((\Box\varphi_t \wedge \mathcal{T} \varphi_c) \to \Box(\Box\varphi_t \wedge \mathcal{T} \varphi_c)) \in \Phi$, they imply that $\exists \Box (\mathcal{T} \varphi_c \wedge \neg \Box \varphi_t) \in \Phi$ by axiom $\exists \mathcal{T} \varphi \wedge \neg \exists \mathcal{T} \psi \wedge \forall((\varphi \wedge \psi) \to \Box(\varphi \wedge \psi)) \to \exists \Box (\varphi \wedge \neg\psi)$. Note that for all $c \in C_1$, this result holds.

We take the union of $\| \Box (\mathcal{T} \varphi_c \wedge \neg \Box \varphi_t) \|$ for each $c \in C_1$, i.e.

$$s = \bigcup_{c \in C_1} \| \Box (\mathcal{T} \varphi_c \wedge \neg \Box \varphi_t) \|$$

and prove two facts about s: $s \cap t = \varnothing$ and for all $c \in C$, $c \cap s \neq \varnothing$. Note that it is "for all $c \in C$" rather than for all $c \in C_1$.

First, $s \cap t = \varnothing$. By $t \subseteq \| \Box \varphi_t \|$ and $\| \Box (\mathcal{T} \varphi_c \wedge \neg \Box \varphi_t) \| \subseteq \| \neg \Box \varphi_t \|$ for all $c \in C_1$, it follows that $t \cap \| \Box (\mathcal{T} \varphi_c \wedge \neg \Box \varphi_t) \| = \varnothing$ for all $c \in C_1$. Therefore, $s \cap t = \varnothing$.

Second, $s \cap c \neq \varnothing$ for all $c \in C$.

For each $c_1 \in C_1$, by theorem $\Box(\mathcal{T}\varphi_{c_1} \wedge \neg\Box\varphi_t) \to \mathcal{T}\varphi_{c_1}$, it follows that $\|\Box(\mathcal{T}\varphi_{c_1} \wedge \neg\Box\varphi_t)\| \cap \|\mathcal{T}\varphi_{c_1}\| \neq \varnothing$. Since $\|\mathcal{T}\varphi_{c_1}\| \subseteq c_1$ for all $c_1 \in C_1$, $s \cap c_1 \neq \varnothing$ for all $c_1 \in C_1$.

Suppose that $s \cap c_2 = \varnothing$ for some $c_2 \in C_2$. Then due to the fact that for all $c_1 \in C_1$, $s \cap c_1 \neq \varnothing$, the supposition makes $c_2 \notin C_2$. Contradiction. Hence $s \cap c_2 \neq \varnothing$ for all $c_2 \in C_2$.

The fact that $s \cap t = \varnothing$ and t is not semi-acceptable tell us that we have found an open s in τ such that $t \twoheadleftarrow s$. $s \cap c \neq \varnothing$ for all $c \in C$ tells us that $s \not\twoheadleftarrow c$ for all $c \in C$. Hence $t \notin d(C)$.

Therefore, we have shown that no matter whether t is semi-acceptable or not, if $t \notin C$, then $t \notin d(C)$ in the second case. Since d is monotonic and $C \subseteq \mathsf{LFP}$ as we have proved, it follows that $d(C) \subseteq d(\mathsf{LFP}) = \mathsf{LFP}$, which implies that $t \notin \mathsf{LFP}$.

Finally, we can conclude that, in both cases, $t \notin C$ implies that $t \notin \mathsf{LFP}$. This completes the proof.

$\Box$

Now we can prove the truth lemma.

Let $[\![\phi]\!]^\Phi = \{\Gamma \in W^\Phi \mid \mathsf{M}^\Phi, \Gamma \vDash \phi\}$.

A.2.12. Lemma. *In the quasi-a-model M^Φ, for all $\phi \in \mathcal{L}_{\forall\mathcal{T}\Box}$ and all $\Gamma \in W^\Phi$,*

$$\Gamma \in \|\phi\|^\Phi \textit{ if and only if } \Gamma \in [\![\phi]\!]^\Phi$$

Proof:

The proofs for the cases of atomic propositions, boolean connectives and $\Box$ and $\forall$ are routine. So we skip them and focus on the case of $\mathcal{T}\phi$.

Assume $\Gamma \in \|\mathcal{T}\varphi\|^\Phi$, which implies that $\exists\mathcal{T}\varphi \in \Phi$ by Proposition A.2.2. By Fact A.2.6, $\exists\mathcal{T}\varphi \in \Phi$ implies that $\|\mathcal{T}\varphi\| \in \tau$. Let $t = \|\mathcal{T}\varphi\|$. Then $\|\mathcal{T}\varphi\| \subseteq t$ and $\exists\mathcal{T}\varphi \in \Phi$ together imply that $t \in C_1$. By Lemma A.2.11, it implies $t \in \mathsf{LFP}$. By axiom $\mathcal{T}\varphi \to \varphi$, $t \subseteq \|\varphi\|$. By inductive hypothesis, $\|\varphi\| = [\![\varphi]\!]$. By the truth condition of $\mathcal{T}$, $\Gamma \in [\![\mathcal{T}\varphi]\!]$.

Assume $\Gamma \in [\![\mathcal{T}\varphi]\!]$. Then by the truth condition of $\mathcal{T}\varphi$, there is $t \in \mathsf{LFP}^\Phi$ such that $\Gamma \in t \subseteq [\![\varphi]\!]$. According to the truth condition of $\Box\varphi$ and the fact that if $\Delta \in t$ then $\geqslant_\Delta \subseteq t$, we have $t \subseteq [\![\Box\varphi]\!]$.

By inductive hypothesis, it can be proved that $[\![\Box\varphi]\!] = \|\Box\varphi\|$. So $\Gamma \in t \subseteq \|\Box\varphi\|$.

If $t \in C_1$, then there exists $\|\mathcal{T}\psi\| \subseteq t$ such that $\exists\mathcal{T}\psi \in \Phi$. Since $\|\mathcal{T}\psi\| \subseteq t \subseteq \|\Box\varphi\|$, then $\forall(\mathcal{T}\psi \to \Box\varphi) \in \Phi$. Take an arbitrary $\Delta \in \|\mathcal{T}\psi\|$. By $\forall(\mathcal{T}\psi \to \Box\varphi) \in \Phi$, we have $\forall(\mathcal{T}\psi \to \Box\varphi) \in \Delta$. Together with theorem $\forall(\varphi \to \psi) \to (\mathcal{T}\varphi \to \mathcal{T}\psi)$ proved in Fact 2.5.6, it implies $\mathcal{T}\mathcal{T}\psi \to \mathcal{T}\Box\varphi \in \Delta$. Since $\Delta \in \|\mathcal{T}\psi\|$, we have $\Delta \in \|\mathcal{T}\mathcal{T}\psi\|$, which implies that $\Delta \in \|\mathcal{T}\Box\varphi\|$, i.e., $\mathcal{T}\Box\varphi \in \Delta$. By axiom $\mathcal{T}\varphi \to \forall(\Box\varphi \to \mathcal{T}\varphi)$, theorem $\mathcal{T}\varphi \leftrightarrow \mathcal{T}\Box\varphi$ and $\mathcal{T}\Box\varphi \in \Delta$, it follows that $\forall(\Box\varphi \to \mathcal{T}\varphi) \in \Delta$. So $\forall(\Box\varphi \to \mathcal{T}\varphi) \in \Phi$, which implies that $\|\Box\varphi\| \subseteq \|\mathcal{T}\varphi\|$. Since $\Gamma \in t \subseteq \|\Box\varphi\|$, we have $\Gamma \in \|\mathcal{T}\varphi\|$.

If $t \in C_2$, then for all $\psi \in \mathcal{L}_{\forall \mathcal{T} \Box}$ such that $t \subseteq \| \Box \psi\|$, there is $\chi \in \mathcal{L}_{\forall \mathcal{T} \Box}$ such that $\| \mathcal{T} \chi\| \subseteq \| \Box \psi\|$ and $\exists \mathcal{T} \chi \in \Phi$. Now since we have $t \subseteq \| \Box \varphi\|$, there exists $\eta \in \mathcal{L}_{\forall \mathcal{T} \Box}$ such that $\| \mathcal{T} \eta\| \subseteq \| \Box \varphi\|$ and $\exists \mathcal{T} \eta \in \Phi$. By copying the argument we use in the case of $t \in C_1$, we can also achieve the result that $\Gamma \in \| \mathcal{T} \varphi\|$.

Therefore, we have proved that $\Gamma \in \| \mathcal{T} \varphi\|$ given that $\Gamma \in [\![\mathcal{T} \varphi]\!]$. So together with $\| \mathcal{T} \varphi\| \subseteq [\![\mathcal{T} \varphi]\!]$, we have proved that $\| \mathcal{T} \varphi\| = [\![\mathcal{T} \varphi]\!]$. This completes the proof.

□

Finally, by Proposition 2.5.12, we only need to prove that M^Φ in Definition A.2.1 is Alexandroff, which will give us a topological argumentation model where Φ_0 is satisfiable by Lemma A.2.12.

A.2.13. Lemma. M^Φ *is Alexandroff.*

Proof:
First observe that whether M^Φ is Alexandroff has nothing to do with $\leftarrowtail$. So we can just apply Proposition 5.6.15 in Özgün (2017), which tells us that if $\tau^\Phi = \{\bigcup_{\Gamma \in S} \geqslant_\Gamma \mid S \subseteq W^\Phi\}$ then M^Φ is Alexandroff. By Fact A.2.5 and $E_0 = \{\bigcup_{\Gamma \in S} \geqslant_\Gamma \mid S \subseteq W^\Phi\} \setminus \{\varnothing\}$, we get $\tau^\Phi = \{\bigcup_{\Gamma \in S} \geqslant_\Gamma \mid S \subseteq W^\Phi\}$. Thus it follows that M^Φ is Alexandroff. □

Appendix B

Appendix of Chapter Three

B.1 Semantic Proof of Proposition 3.4.2 and Theorem 3.4.6

B.1.1. LEMMA. *Given a topological argumentation model $\mathcal{M}$ for all $f \in \mathsf{LFP}_\tau$, if there are $u, v \in f$ such that $v \sqsubset_\tau u$, then there is $f' \in \mathsf{LFP}_\tau$ such that $f' \subset f$ and $v \notin f'$ and $u \in f'$.*

Proof:

Take an arbitrary argument $\mathsf{f} \in \mathsf{LFP}_\tau$. Assume there are $u, v \in \mathsf{f}$ such that $v \sqsubset_\tau u$. We first prove that for all $x \in \tau$ such that

a. $x \subset \mathsf{f}$ and

b. $u \in x$ and $v \notin x$

c. $x \notin \mathsf{LFP}_\tau$

there is $y \in \tau$ such that

A. $x \cap y = \varnothing$

B. $y \cap \mathsf{f} \neq \varnothing$

C. $v \notin y \cap \mathsf{f}$

Take an arbitrary $x \in \tau$ which satisfies conditions a, b and c. By c, there is $y \in \tau$ such that $x \twoheadleftarrow y$ and there is no $f \in \mathsf{LFP}_\tau$ such that $y \twoheadleftarrow f$. $x \twoheadleftarrow y$ implies that $x \cap y \neq \varnothing$(A) ; $y \cap f \neq \varnothing$ for all $f \in \mathsf{LFP}_\tau$ follows from the fact that there is no $f \in \mathsf{LFP}_\tau$ such that $y \twoheadleftarrow f$. So $y \cap \mathsf{f} \neq \varnothing$ (B). $x \cap y = \varnothing$ implies that $x \cap (y \cap \mathsf{f}) = \varnothing$, which together with $u \in x$ implies that $u \notin y \cap \mathsf{f}$. $u \notin y \cap \mathsf{f}$, together with $v \sqsubset_\tau u$, implies that $v \notin y \cap \mathsf{f}$ (C).

Next, we prove that there must be an argument satisfying all the requirements in the Lemma.

For each $x \in \tau$ which satisfy conditions a, b and c, let

$$C_x = \{y \cap \mathsf{f} \mid y \in \tau \text{ and it satisfies conditions A, B and C}\} \ .$$

Note that $\bigcup C_x \in \tau$ and $x \cap \bigcup C_x = \varnothing$ and $\bigcup C_x \subset \mathsf{f}$ and $v \notin \bigcup C_x$.

We then define a partial function $H : \tau \mapsto \tau$ such that given any $x \in \tau$, if it satisfies conditions a, b and c, then

$$H(x) = x \cup \bigcup C_x$$

else if it only satisfies conditions a and b but not c, then

$$H(x) = x$$

else $H(x)$ is undefined.

Observe that if $x \in \tau$ satisfies conditions a and b, then $H(x) \in \tau$ and satisfies conditions a and b. More generally speaking, for all sets $X \subseteq \tau$, if every $x \in X$ satisfies conditions a and b, then $\bigcup X \in \tau$ and satisfies condition a and b.

By the assumption that there are $u, v \in \mathsf{f} \in \mathsf{LFP}_\tau$ such that $v \sqsubset_\tau u$, it follows that there is $t \in \tau$ such that $u \in t$ but $v \notin t$.

Thus we can define inductively the following *non-decreasing* sequence by taking an arbitrary $t \in \tau$ such that $u \in t$ but $v \notin t$.

$$\begin{aligned} t_0 &= t \cap \mathsf{f} \\ t_{\alpha+1} &= H(t_\alpha) \\ t_\beta &= \bigcup_{\alpha<\beta} t_\alpha \qquad \text{whenever } \beta \text{ is a limit ordinal} \end{aligned}$$

Take the set $U = \{t_\alpha \mid \alpha \text{ is an ordinal}\}$. Because it is a well-founded ordered subset of 2^W, there is an ordinal β and a *monotonic* bijection r from U onto the set $\{\alpha \mid \alpha < \beta\}$. Let $r' : \{\alpha \leq \beta\} \mapsto U$ be the *monotonic* mapping defined by $r'(\alpha) = t_\alpha$. If it were injective, then the *monotonic* mapping $r \circ r' : \{\alpha < \beta\} \mapsto \{\alpha \leq \beta\}$ would be injective too, which is impossible. Thus there exists γ and γ' such that $\gamma < \gamma + 1 \leq \gamma' \leq \beta$ and $r'(\gamma) = r'(\gamma')$. This implies that $t_{\gamma+1} = H(t_\gamma) = t_\gamma$.

Since $H(t_\gamma) = t_\gamma$ and $\bigcup C_x \nsubseteq x$ for all $x \in \tau$, t_γ satisfies conditions a and b but not c by the definition of the function H. That is, (a) $t_\gamma \subset \mathsf{f}$, (b) $u \in t_\gamma$ and $v \notin t_\gamma$ and not (c) $t_\gamma \in \mathsf{LFP}_\tau$. □

Proof of Proposition 3.4.2
Proof:
Take any $m \in \mathsf{Mini}$. By Lemma B.1.1, there cannot be $v, u \in m$ such that $v \sqsubset_\tau u$. So $m \subseteq P$ if and only if $m \rightsquigarrow P$ holds in $\mathcal{M}$. □

Proof of Theorem 3.4.6

B.1.2. LEMMA. *Given a topological argumentation model $\mathcal{M}$ whose topological space is an Alexandroff space, take $\sqsubseteq_\tau$ as the agent's normality order. Then, for all $f \in \mathsf{LFP}_\tau$, if $f \rightsquigarrow P$ holds in $\mathcal{M}$, then there is $f' \subseteq f$ such that $f' \in \mathsf{LFP}_\tau$ and $f' \subseteq P$.*

Proof:
Take an arbitrary $\mathsf{f} \in \mathsf{LFP}_\tau$ and assume that $\mathsf{f} \nsubseteq P$.

We **first** prove that there is $t \in \tau$ such that $t \neq \varnothing$ and $t \subseteq \mathsf{f} \cap P$. Since $\mathsf{f} \nsubseteq P$, $\mathsf{f} \cap \overline{P} \neq \varnothing$. $\mathsf{f} \cap P \neq \varnothing$ follows from the fact that $\mathsf{f} \rightsquigarrow P$ holds.

Let $u \uparrow = \{w \in W \mid u \sqsubseteq_\tau w\}$. For all $v \in \mathsf{f} \cap \overline{P}$, there is $u \in \mathsf{f} \cap P$ such that $v \sqsubseteq_\tau u$ and $u \uparrow \subseteq P$ by the fact that $\mathsf{f} \rightsquigarrow P$ holds in $\mathcal{M}$. By $v \sqsubseteq_\tau u$ and $u \uparrow \subseteq P$, it follows that $v \sqsubset_\tau u$. We take such an argument u for each $v \in \mathsf{f} \cap \overline{P}$ and denote it by u_v.

Since $v \sqsubset_\tau u_v$ and $v, u_v \in \mathsf{f}$, by Lemma B.1.1, there is $f \in \mathsf{LFP}_\tau$ for $v \in \mathsf{f} \cap \overline{P}$ such that $f \subset \mathsf{f}$ and $v \notin f$ and $u \in f$. We take such an argument f for each $v \in \mathsf{f} \cap \overline{P}$ and denote it by f_v.

Since $u_v \in f_v$, it follows that $u_v \uparrow \subseteq f_v$. Because $u_v \uparrow \subseteq P$, for all $v' \in \mathsf{f} \cap \overline{P}$ such that $v' \in f_v$, we have $u_v \not\sqsubseteq_\tau v'$. Hence for all $v' \in \mathsf{f} \cap \overline{P}$ such that $v' \in f_v$, there is a $t_{v'} \in \tau$ such that $u_v \in t_{v'}$ and $v' \notin t_{v'}$. We take such an open for each $v' \in \mathsf{f} \cap \overline{P}$ such that $v' \in f_v$ and denote it by $t_{v'}$. Note that $u_v \uparrow \subseteq t_{v'}$.

To sum up, we have for each $v \in \mathsf{f} \cap \overline{P}$ two types of opens u_v and f_v. u_v satisfies $u_v \in \mathsf{f} \cap P$ and $u_v \uparrow \subseteq P$ and $v \sqsubset_\tau u_v$; f_v satisfies $f_v \subseteq \mathsf{f}$ and $u_v \uparrow \subseteq f_v$ and $v \notin f_v$. For each $v' \in \mathsf{f} \cap \overline{P} \cap f_v$ we have one type of opens $t_{v'}$. $t_{v'}$ satisfies $u_v \uparrow \subseteq t_{v'}$ and $v' \notin t_{v'}$.

Take $t_{u_v} = f_v \cap \bigcap_{v' \in \mathsf{f} \cap \overline{P} \cap f_v} t_{v'}$. By $u_v \uparrow \subseteq t_{v'}$ for all $v' \in \mathsf{f} \cap \overline{P} \cap f_v$ and $u_v \uparrow \subseteq f_v$, it follows that $t_{u_v} \neq \varnothing$. By $v' \notin t_{v'}$ for all $v' \in \mathsf{f} \cap \overline{P} \cap f_v$, it follows that $v' \notin t_{u_v}$. By $f_v \subseteq \mathsf{f}$ for all $v \in \mathsf{f} \cap \overline{P}$, it follows that $\mathsf{f} \cap \overline{P} \cap f_v = \overline{P} \cap f_v$. So $v' \notin t_{u_v}$ for all $v' \in \overline{P} \cap f_v$, which, together with the fact that $t_{u_v} \subseteq f_v$, implies that $t_{u_v} \subseteq P$ for all $v \in \mathsf{f} \cap \overline{P}$.

Hence for each $v \in \mathsf{f} \cap \overline{P}$, we can find an open t_{u_v} (the topological space is an Alexandroff space) such that $t_{u_v} \neq \varnothing$ and $t_{u_v} \subseteq P \cap \mathsf{f}$.

Hence we have proved that there is $t \in \tau$ such that $t \neq \varnothing$ and $t \subseteq \mathsf{f} \cap P$.

Next take $t_p = \bigcup\{t \in \tau \mid t \subseteq \mathsf{f} \cap P\}$. Note that $t_p \neq \varnothing$ and $t_p \subseteq \mathsf{f} \cap P \subset \mathsf{f}$ (by $\mathsf{f} \nsubseteq P$). We prove that $t_p \in \mathsf{LFP}_\tau$. Suppose not. Then there is $t' \in \tau$ such that $t_p \twoheadleftarrow t'$ and for all $f \in \mathsf{LFP}_\tau$, $f \cap t' \neq \varnothing$. So $\mathsf{f} \cap t' \neq \varnothing$. By $t_p \cap t' = \varnothing$, which follows from $t_p \twoheadleftarrow t'$, it follows that $t' \nsubseteq \mathsf{f} \cap P$. By $\mathsf{f} \cap t' \neq \varnothing$ and $t' \nsubseteq \mathsf{f} \cap P$, it follows that $t' \cap \mathsf{f} \cap \overline{P} \neq \varnothing$.

Since $t' \cap t_p = \varnothing$, there is no $t \in \tau$ such that $t \subseteq t' \cap \mathsf{f} \cap P$. So for all $t \subseteq t' \cap \mathsf{f}$ which is non-empty, there is $x \in t$ such that $x \in t' \cap \mathsf{f} \cap \overline{P}$.

For each $w \in t' \cap \mathsf{f}$, take $t_w = \bigcap_{w \in t \in \tau} t$. Since the topological space is an Alexandroff space, t_w is an open. Since $w \in t_w$, t_w is non-empty. Moreover, $w \in t' \cap \mathsf{f} \in \tau$ implies that $t_w \subseteq t' \cap \mathsf{f}$. By the conclusion in the previous paragraph, there is $x \in t_w$ such that $x \in t' \cap \mathsf{f} \cap \overline{P}$. $x \in t_w$ implies that $w \sqsubseteq_\tau x$. So for each

$w \in t' \cap \mathsf{f}$, we find a $x \in \overline{P}$ such that $w \sqsubseteq_\tau x$. This contradicts the assumption that $\mathsf{f} \rightsquigarrow P$ holds in $\mathcal{M}$, because $w \uparrow\in t' \cap \mathsf{f}$.

$\square$

Appendix C

Appendix of Chapter Four

C.1 Proof of Lemma 4.2.12

Proof:
(⇒) The Knaster-Tarski fixpoint theorem (Knaster, 1928; Tarski, 1955) states that GFP^P is the union of all d^P's postfix points, $\mathsf{GFP}^P = \bigcup\{X \subseteq \mathsf{Ag} \mid X \subseteq d^P(X)\}$. Hence, as any P-admissible set A is, by definition, a postfix point of d^P, we have $A \subseteq \mathsf{GFP}^P$ and thus $s \in A$ implies $s \in \mathsf{GFP}^P$.

(⇐) Take any $s \in \mathsf{GFP}^P$ and define $A := \mathsf{Gfp}.d \cap \mathsf{Def}(s)$; it is enough to show that A is a P-admissible set containing s, i.e., 1) A is conflict free with respect to $\twoheadleftarrow^P$ and 2) $A \subseteq d^P(A)$ and 3) $s \in A$.

For the first, observe that any two arguments $x, y \in A$ are, by definition, also in $\mathsf{Def}^P(s)$. Thus, each one of them is at a distance of even steps from s and hence cannot attack each other (as otherwise, the attacker would be at a distance odd steps from s, which is impossible as $\twoheadleftarrow^P$ is uncontroversial). Therefore, A is conflict-free.

For the second, take any argument x attacking some $s' \in A$. Since $s' \in \mathsf{GFP}^P$, there must be $s'' \in \mathsf{GFP}^P$ such that $x \twoheadleftarrow^P s''$ (recall, GFP^P is a fixed point of d^P, so $\mathsf{GFP}^P = d^P(\mathsf{GFP}^P)$); hence, $s'' \in \mathsf{Def}^P(s')$. But then, as $s' \in \mathsf{Def}^P(s)$ (by A's definition) and $s'' \in \mathsf{Def}^P(s')$, it follows that $s'' \in \mathsf{Def}^P(s)$, that is, $s'' \in A$ (by A's definition). Hence, every argument x attacking some argument (s') in A is in turn attacked by an argument (s'') that belongs to A; thus, $A \subseteq d^P(A)$.

For the third, since clearly $s \in \mathsf{Def}^P(s)$, $s \in \mathsf{GFP}^P$ implies $s \in A$. □

C.2 Proof of Theorem 4.3.4

In this section we show that for all $\varphi \in \mathcal{L}_{\alpha\beta}$, $\vDash \varphi$ implies $\vdash \varphi$. We first define the partial multi-agent argumentation model and then extend it to a multi-agent argumentation model.

C.2.1. Definition (Partial MAM). A partial multi-agent argumentation model $\mathcal{M}^X$ is a structure $(W, \mathsf{Ag}, \{\twoheadleftarrow^P\}^{P\in X\subseteq 2^W}, f, V)$ where $W, \mathsf{Ag}, \twoheadleftarrow^P, f$ and V are defined as in the multi-agent argumentation model, and X is closed under complement, i.e. if $P \in X$, then $\overline{P} \in X$. Moreover, it satisfies a restricted version of the conditions we impose on the multi-agent argumentation model:

1. For all $P \in X$, $s_1 \twoheadleftarrow^P s_2$ if and only if $s_1 \twoheadleftarrow^{\overline{P}} s_2$.
2. For all $P \in X$, if $s_1 \twoheadleftarrow^P s_2$,
   (a) either $f(s_1) \subseteq P$ or $f(s_1) \subseteq \overline{P}$; and
   (b) $f(s_1) \subseteq P$ implies $f(s_2) \subseteq \overline{P}$.
3. For all $P, Q \in X$, if $s_1 \twoheadleftarrow^P s_2$ and $f(s_1) \subseteq Q \subseteq P$, then $s_1 \twoheadleftarrow^Q s_2$.

In words, a partial MAM $\mathcal{M}^X$ is simply a MAM that only 'discusses' propositions in X, and therefore only needs to satisfy the frame conditions relative to the elements in X. All MAM are partial MAMs; for the other direction, we have the following lemma.

C.2.2. Lemma. *Any partial MAM $\mathcal{M}^X$ can be extended to a MAM.*

Proof:
We define the relation $\twoheadleftarrow^P$ for each $P \notin X$ as follows:

$$s \twoheadleftarrow^P s' \quad \text{if and only if} \quad \begin{array}{l} \text{there is a set } Q \in X \text{ such that } s \twoheadleftarrow^Q s' \text{ and} \\ f(s) \subseteq P \subseteq Q \text{ or } f(s) \subseteq \overline{P} \subseteq Q \end{array}$$

We claim that the model generated by adding these attacking relations into $\mathcal{M}^X$ is a MAM. To prove this claim, we only need to check that it satisfies the four frame conditions.

1: $s_1 \twoheadleftarrow^P s_2$ if and only if $s_1 \twoheadleftarrow^{\overline{P}} s_2$. We only need to prove that if $s_1 \twoheadleftarrow^P s_2$, then $s_1 \twoheadleftarrow^{\overline{P}} s_2$.

- If $P \in X$, it follows immediately from the restricted version of condition 1 and the fact that X is closed under complement.
- If $P \notin X$, the definition of $s_1 \twoheadleftarrow^P s_2$ also implies that $s_1 \twoheadleftarrow^{\overline{P}} s_2$, since the definition is symmetric.

2(a): if $s_1 \twoheadleftarrow^P s_2$, either $f(s_1) \subseteq P$ or $f(s_1) \subseteq \overline{P}$. For $P \in X$, by the restricted version of condition 2(a), $\twoheadleftarrow^P$ satisfies condition 2(a). For $P \notin X$, the definition of $\twoheadleftarrow^P$ implies $f(s_1) \subseteq P$ or $f(s_1) \subseteq \overline{P}$

2(b): if $s_1 \twoheadleftarrow^P s_2$ and $f(s_1) \subseteq P$, then $f(s_2) \subseteq \overline{P}$. For $P \in X$, by the restricted version of condition 2(b), $\twoheadleftarrow^P$ satisfies condition 2(b). For $P \notin X$, the definition of $\twoheadleftarrow^P$ implies that there is $Q \in X$ such that $f(s_1) \subseteq P \subseteq Q$. By 2(b) for $Q \in X$, $f(s_2) \subseteq \overline{Q} \subseteq \overline{P}$.

3: if $s_1 \twoheadleftarrow^P s_2$ and $f(s_1) \subseteq Q \subseteq P$, then $s_1 \twoheadleftarrow^Q s_2$.

- If $P, Q \in X$, this condition follows from its restricted version of condition 3.
- If $P \in X$ and $Q \notin X$, take P as the set required by the definition of $\twoheadleftarrow^Q$, it follows immediately that $s_1 \twoheadleftarrow^Q s_2$
- If $P \notin X$ and $Q \in X$, by definition of $\twoheadleftarrow^P$, there is a set $R \in X$ such that $s_1 \twoheadleftarrow^R s_2$ and $f(s_1) \subseteq P \subseteq R$ or $f(s_1) \subseteq \overline{P} \subseteq R$. (Note that the second case is not possible since $f(s_1) \subseteq P$ by assumption.) By the restricted version of condition 3, since $Q \in X$ and $f(s_1) \subseteq Q \subseteq P \subseteq R$, we have $s_1 \twoheadleftarrow^Q s_2$.
- If $P, Q \notin X$, by the definition of $\twoheadleftarrow^P$, there is a set $R \in X$ such that $s_1 \twoheadleftarrow^R s_2$ and $f(s_1) \subseteq P \subseteq R$ or $f(s) \subseteq \overline{P} \subseteq R$. Take the set R, since $f(s_1) \subseteq Q \subseteq P \subseteq R$, it follows that $s_1 \twoheadleftarrow^Q s_2$ by the definition of $s_1 \twoheadleftarrow^Q s_2$.

□

Hence, for an arbitrary GBAS-consistent formula $\varphi \in \mathcal{L}_{\alpha\beta}$, we first build a partial MAM in which it is satisfied by using maximal consistent sets. When doing so, we take care of the fact that there are two dimensions in a MAM. Corresponding to these two dimensions, the language $\mathcal{L}_{\alpha\beta}$ is divided into two parts – α-formulas and β-formulas. Note that given an GBAS-consistent set of α-formulas A and an GBAS-consistent set of β-formulas B, their union $A \cup B$ is GBAS-consistent; (suppose otherwise, then there must be $\alpha_1, \ldots, \alpha_n$ and $\beta_1, \ldots, \beta_m$ such that $\vdash \bigwedge_{i=1}^n \alpha_i \to \neg \bigwedge_{i=1}^m \beta_i$. However, this is impossible since $\bigwedge_{i=1}^n \alpha_i \to \neg \bigwedge_{i=1}^m \beta_i$ is not a formula in $\mathcal{L}_{\alpha\beta}$.) This fact gives us the flexibility to construct the pairs of possible worlds and arguments on which we evaluate formulas.

Given a formula φ, we define $\sim$ in the following formula:

$$\sim \varphi := \begin{cases} \psi & \text{if } \varphi \text{ is of the form } \neg\psi, \\ \neg\varphi & \textit{otherwise} \end{cases}$$

A set of formulas $X \subseteq \mathcal{L}_{\alpha\beta}$ is closed under single negation if and only if $\sim \varphi$ belongs to X whenever $\varphi \in X$.

C.2.3. Definition. Let X be a set of formulas. The set X is *FL-closed* if and only if it is closed under subformulas (e.g., if $[\alpha]\beta \in X$, then $\alpha, \beta \in X$) and it satisfies the following additional constraints:

- $\top, \boxminus\bot, \boxbar\bot, \Box\bot \in X$
- if $[\alpha]\beta \in X$, then $[\sim \alpha]\beta, \Box\alpha, \langle\alpha\rangle\top \in X$;
- if $\mathsf{Gfp}^{\alpha} \in X$, then $[\alpha]\langle\alpha\rangle\,\mathsf{Gfp}^{\alpha} \in X$.

Next, we build the canonical model starting by constructing maximal consistent sets. Given the two kinds of formulas, α-formulas and β-formulas, intricate coordination between them is required during the construction.

C.2.4. Definition. Let Σ be a set of formulas. We define $Sub(\Sigma)$ as the smallest set containing Σ which is FL-closed and closed under single negations. And $Sub^{+}(\Sigma)$ is the smallest set containing $Sub(\Sigma)$ which satisfies the following two conditions:

- if $\boxminus\alpha \in Sub(\Sigma)$, then $\boxbar\,\Box\,\boxminus\alpha \in Sub^{+}(\Sigma)$ and $\boxbar\,\Box\,\neg\,\boxminus\,\alpha \in Sub^{+}(\Sigma)$;
- if $\boxbar\beta \in Sub(\Sigma)$, then $\boxbar\,\Box\,\boxbar\beta \in Sub^{+}(\Sigma)$ and $\boxbar\,\Box\,\neg\,\boxbar\,\beta \in Sub^{+}(\Sigma)$;

The set $Cls(\Sigma)$, the closure of Σ, is the the smallest set containing $Sub^{+}(\Sigma)$ which is FL-closed and closed under single negations.

C.2.5. Definition. (Atoms) Let Σ be a set of formulas. And let $\mathcal{L}_{\alpha} \subseteq \mathcal{L}_{\alpha\beta}$ be the set of all α-formulas and $\mathcal{L}_{\beta} \subseteq \mathcal{L}_{\alpha\beta}$ be the set of all β-formulas

- A set of formulas Γ is an atom over Σ if it is a maximal consistent subsets of $Cls(\Sigma)$. The set $At(\Sigma)$ contains all the atoms over Σ.
- A set of formulas A is an α-atom over Σ if it is a maximal consistent subsets of $Cls(\Sigma) \cap \mathcal{L}_{\alpha}$. The set $At^{\alpha}(\Sigma)$ contains all α-atoms over Σ.
- A set of formulas B is an β-atom over Σ if it is a maximal consistent subsets of $Cls(\Sigma) \cap \mathcal{L}_{\beta}$. The set $At^{\beta}(\Sigma)$ contains all β-atoms over Σ.

C.2.6. Fact.

- If A is an α-atom over Σ and B is a β-atom over Σ, then $A \cup B$ is an atom over Σ;
- If Γ is an atom over Σ, then $\Gamma \cap \mathcal{L}_{\alpha}$ is an α-atom and $\Gamma \cap \mathcal{L}_{\beta}$ is a β-atom.

Proof:
For the first, we know that $A \cup B \subseteq Cls(\Sigma)$ is GBAS-consistent; thus, we only need to prove that any X satisfying $A \cup B \subset X \subseteq Cls(\Sigma)$ is inconsistent. This is obvious, since any $\psi \in Cls(\Sigma)$ with $\psi \notin A \cup B$ is either an α-formula or β-formula. Without loss of generality we can assume it is an α-formula; since A is

an α-atom over Σ, the set $A \cup \{\psi\}$ must be GBAS-inconsistent. Thus, X is also GBAS-inconsistent.

For the second, suppose $\Gamma \cap \mathcal{L}_\alpha = A$ is not an α-atom. Since A is consistent, there is an α-formula α' not in A such that $A \cup \{\alpha'\}$ is still consistent, and thus $\Gamma \cup \{\alpha'\}$ is also consistent. But $\alpha' \notin A$ implies $\alpha' \notin \Gamma$, and thus this contradicts the maximality of Γ. Therefore, $\Gamma \cap \mathcal{L}_\alpha$ must be an α-atom. The argument for $\Gamma \cap \mathcal{L}_\beta$ is similar. $\square$

C.2.7. Lemma. *If $\Phi \subseteq Cls(\Sigma)$ and Φ is consistent, then there is a $\Gamma \in At(\Sigma)$ such that $\Phi \subseteq \Gamma$.*

The proof of Lemma C.2.7, analogous to the Lindenbaum's Lemma, follows the same argument as the proof of Lemma 4.83 in Blackburn, Rijke and Venema (2002). Together with the second fact of Fact C.2.6, it implies the following lemmas:

C.2.8. Lemma.

1. *If $X \subseteq Cls(\Sigma) \cap \mathcal{L}_\alpha$ is consistent, then there is a $A \in At^\alpha(\Sigma)$ such that $X \subseteq A$.*
2. *If $Y \subseteq Cls(\Sigma) \cap \mathcal{L}_\beta$ is consistent, then there is a $B \in At^\beta(\Sigma)$ such that $Y \subseteq B$.*

We can now fix a formula $\phi \in \mathcal{L}$ and construct the canonical model for it, which will be proved to be a partial MAM. Before doing so here is first some useful notations.

Let X and Σ be sets of formulas.

- If X is finite, define $\widehat{X} := \bigwedge_{\varphi \in X} \varphi$.
- For $\circ \in \{\boxminus, \boxbar, \Box, [\alpha]\}$, define the sets

$$X^\circ := \{\varphi \in \mathcal{L} \mid \circ\varphi \in X\}, \qquad \circ X^\circ := \{\circ\varphi \in \mathcal{L} \mid \circ\varphi \in X\},$$

  If X is finite, define

$$X^\Sigma_\circ := \{\varphi \in Cls(\Sigma) \mid \vdash \widehat{X^\circ} \to \varphi\} \enspace .$$

  When Σ is a singleton $\{\sigma\}$, the set $X^{\{\sigma\}}_\circ$ will be abbreviated as $X^\sigma_\circ$.

Second, the following proof shows a property required by the definition of our canonical model.

C.2.9. Lemma. *Given a consistent ϕ, there is $\Delta \in At(\phi)$ such that $\phi \in \Delta$ and $\Delta^\phi_\boxbar \subseteq \Delta$.*

Proof:

The proof is divided into two cases.

First, if ϕ is an α-formula then, by Lemma C.2.8 (a), there is an α-atom A such that $\phi \in A$, due to its consistency. Note that $A^{\boxplus}$ is consistent (otherwise, $\boxplus\bot \in A$, contradicting axiom **D**). Thus, by this and the fact that $A^{\boxplus} \subseteq Cls(\phi)$ (Lemma C.2.8 (b)), there is a β-atom B such that $A^{\boxplus} \subseteq B$, which implies that $A^{\phi}_{\boxplus} \subseteq B$. By Fact C.2.6, $A \cup B \in At(\phi)$.

Second, if ϕ is a β-formula, suppose that there is no $\Delta \in At(\phi)$ such that $\phi \in \Delta$ and $\Delta^{\phi}_{\boxplus} \subseteq \Delta$. Then there is no α-atom A such that $A^{\phi}_{\boxplus} \cup \{\phi\}$ is consistent. (Otherwise, $A^{\phi}_{\boxplus} \cup \{\phi\}$ can be extended to a β-atom by Lemma C.2.8 (b), and by Fact C.2.6 $A \cup B$ is an atom over ϕ: a contradiction.)

Now take the α-formula $\boxplus\phi$. It can be proved that $\boxplus\phi$ is consistent. (Otherwise, $\vdash \neg \boxplus \phi$, which implies that $\vdash \boxplus\phi \leftrightarrow \boxplus\bot$. Since $\vdash \boxplus(\bot \to \neg\phi)$ and $\vdash \boxplus(\bot \to \neg\phi) \wedge \boxplus\bot \to \boxplus\neg\phi$, it follows that $\vdash \boxplus\phi \to \boxplus\neg\phi$, which implies that ϕ is inconsistent; a contradiction.) Since $\boxplus\phi$ is an α-formula, from the first auxiliary result it follows that there is an atom over $\boxplus\phi$, say Γ, such that $\Gamma^{\boxplus\phi}_{\boxplus} \subseteq \Gamma$ and $\boxplus\phi \in \Gamma$.

Next, take $\Delta := \Gamma \cap Cls(\phi)$. It can be shown that Δ is an atom over ϕ (i.e., $\Delta \in At(\phi)$). So take the α-atom $\Theta := \Delta \cap \mathcal{L}_{\alpha}$. Our supposition implies the inconsistency of $\Theta^{\phi}_{\boxplus} \cup \{\phi\}$, as we have shown at the beginning. However, this implies that $\widehat{\boxplus(\Theta^{\phi}_{\boxplus} \cup \{\phi\})}$ is inconsistent. $\boxplus\Theta^{\boxplus} \cup \{\boxplus\phi\} \subseteq \Gamma$ implies that $\vdash \widehat{\Gamma} \to \widehat{\boxplus(\Theta^{\boxplus} \cup \{\phi\})}$. By the definition of $\Theta^{\phi}_{\boxplus}$, $\vdash \widehat{\Theta^{\boxplus}} \to \widehat{\Theta^{\phi}_{\boxplus}}$. Thus, we have $\vdash \widehat{\Gamma} \to \widehat{\boxplus(\Theta^{\phi}_{\boxplus} \cup \{\phi\})}$. Since $\widehat{\boxplus(\Theta^{\phi}_{\boxplus} \cup \{\phi\})}$ is inconsistent, it follows that Γ is inconsistent, contradicting the Γ's consistency. □

Here is, then, the definition of the partial MAM for ϕ.

C.2.10. DEFINITION (Partial MAM over ϕ). Take any $\Delta \in At(\phi)$ satisfying both $\phi \in \Delta$ and $\Delta^{\phi}_{\boxplus} \subseteq \Delta$ (by Lemma C.2.9, such a Δ exists). The *partial MAM over ϕ* is the structure

$$\mathfrak{M}_{\Delta}^{\{\|\alpha\| | [\alpha]\top \in Cls(\phi)\}} = (W, \mathsf{Ag}, f, \{\leftarrow^{\|\alpha\|} | [\alpha]\top \in Cls(\phi)\}, V)^{1}$$

given by

- $W := \{A \in At^{\alpha}(\phi) \mid \boxminus A^{\boxminus} \cap Sub(\Sigma) = \boxminus\Delta^{\boxminus} \cap Sub(\Sigma)$ and $\boxplus A^{\boxplus} \cap Sub(\Sigma) = \boxplus\Delta^{\boxplus} \cap Sub(\Sigma)\}$;
- $\|\alpha\| := \{A \in W \mid \alpha \in A\}$;

[1] Note that $\{\|\alpha\| \mid [\alpha]\top \in Cls(\phi)\}$ will play the role of the labels of attack relations in the partial MAM.

- $\mathsf{Ag} := \{B \in At^{\beta}(\phi) \mid \Delta^{\phi}_{\boxbar} \subseteq B\}$;
- $A \in f(B)$ if and only if $\widehat{B} \wedge \Diamond\widehat{A}$ is consistent;
- for all $\|\alpha\| \in \{\|\alpha\| \mid [\alpha]\top \in Cls(\phi)\}$, $B \twoheadleftarrow^{\|\alpha\|} B'$ if and only if $B \wedge \langle\alpha\rangle B'$ is consistent;
- $V(p) := \{A \in At^{\alpha}(\phi) \mid p \in A\}$

We first prove four existence lemmas for $\boxminus$, $\boxbar$, $\Box$ and $[\alpha]$. During the proof, we will use the following abbreviations:

$$DY := \{\boxminus\alpha \in Sub(\phi) \mid \boxminus\alpha \in \Delta\}, \qquad DN := \{\neg \boxminus \alpha \in Sub(\phi) \mid \neg \boxminus \alpha \in \Delta\},$$
$$EY := \{\boxbar\beta \in Sub(\phi) \mid \boxbar\beta \in \Delta\}, \qquad EN := \{\neg \boxbar \beta \in Sub(\phi) \mid \neg \boxbar \beta \in \Delta\},.$$
$$YN := DY \cup DN \cup EY \cup EN.$$

C.2.11. LEMMA. *In the partial MAM over ϕ, for all $\boxminus\alpha \in Cls(\phi)$, $\boxminus\alpha \notin A \in W$ if and only if there is an α-atom $A' \in W$ such that $\neg\alpha \in A'$.*

Proof:
Assume that $\boxminus\alpha \notin A$ where $\boxminus\alpha \in Cls(\phi)$. We show that $\{\neg\alpha\} \cup YN$ is consistent. Suppose not; then $\vdash \widehat{YN} \to \alpha$. By applying the necessitation rule, we have $\vdash \boxminus\widehat{YN} \to \boxminus\alpha$. By axiom **4** and **5** for $\boxminus$ and $A^{\boxminus} = \Delta^{\boxminus}$, we have $\vdash \widehat{A} \to \boxminus(\widehat{DY} \wedge \widehat{DN})$. By Axioms $\boxbar$**1**, $\boxbar$**2** and $A^{\boxbar} = \Delta^{\boxbar}$, we have $\vdash \widehat{A} \to \boxminus(\widehat{EY} \wedge \widehat{EN})$. Together with $\vdash \boxminus\widehat{YN} \to \boxminus\alpha$, it implies $\boxminus\alpha \in A$, a contradiction. Therefore, we can extend $\{\neg\alpha\} \cup YN$ to an α-atom over ϕ, A', such that $\neg\alpha \in A'$. Moreover, $A'^{\boxminus} = \Delta^{\boxminus}$ and $A'^{\boxbar} = \Delta^{\boxbar}$, so $A' \in W$.

For the other direction, assume there is a α-atom $A' \in W$ such that $\neg\alpha \in A'$. It follows immediately that $\boxminus\alpha \notin A$. (Otherwise, by $A^{\boxminus} = \Delta^{\boxminus} = A'^{\boxminus}$ and axiom **T** for $\boxminus$, we have $\alpha \in A'$, which contradicts the assumption $\neg\alpha \in A'$.) $\Box$

C.2.12. LEMMA. *In the partial MAM over ϕ, $\Delta^{\phi}_{\boxbar} = \bigcap \mathsf{Ag}$*

Proof:
As the $\subseteq$ direction is obvious, we focus on the $\supseteq$ direction. Assume $\beta \in Cls(\phi)$ but $\beta \notin \Delta^{\phi}_{\boxbar}$; we will show that there is a $B \in \mathsf{Ag}$ such that $\beta \notin B$. This amounts to show that $\{\neg\beta\} \cup \Delta^{\phi}_{\boxbar}$ is consistent. Suppose not; then $\vdash \widehat{\Delta^{\phi}_{\boxbar}} \to \beta$. Since for each $\beta \in \Delta^{\phi}_{\boxbar}$, we have $\vdash \widehat{\Delta^{\boxbar}} \to \beta$. So $\vdash \widehat{\Delta^{\boxbar}} \to \widehat{\Delta^{\phi}_{\boxbar}}$. By $\vdash \widehat{\Delta^{\phi}_{\boxbar}} \to \beta$, it follows that $\vdash \widehat{\Delta^{\boxbar}} \to \beta$, which implies that $\beta \in \Delta^{\phi}_{\boxbar}$ by the definition of $\Delta^{\phi}_{\boxbar}$. However, this is contradictory to the assumption that $\beta \notin \Delta^{\phi}_{\boxbar}$. So there must be a $B \in \mathsf{Ag}$ such that $\beta \notin B$. Therefore, $\beta \notin \bigcap \mathsf{Ag}$. $\Box$

C.2.13. Lemma (Existence Lemma for α-formulas). *In the partial MAM over ϕ, for all $\Box\alpha \in Cls(\phi)$, $\Box\alpha \notin B$ if and only if there is an α-atom $A \in W$ such that there is $A \in f(B)$ such that $\neg\alpha \in A$.*

Proof:
Assume $\Box\alpha \notin B$; we will show that $\{\neg\alpha\}$ can be extended to an α-atom $A \in W$ such that $\widehat{B} \wedge \Diamond\widehat{A}$ is consistent. For this we follow the argument of Blackburn, Rijke and Venema (2002, Lemma 4.86) and construct an appropriate α-atom A by forcing choices. So, enumerate the α-formulas in $Cls(\phi)$ as $\alpha_1, \ldots, \alpha_m$, and define A_0 as $\{\neg\alpha\} \cup YN$, with YN as above.

We first prove that $\widehat{B} \wedge \Diamond\widehat{A_0}$ is consistent. Suppose otherwise; then $\vdash \widehat{B} \to \Box\neg\widehat{A_0}$ and $\neg\widehat{A_0} = \alpha \vee \neg\widehat{YN}$. Thus, $\vdash \widehat{B} \to \Box(\alpha \vee \neg\widehat{YN})$, which implies that $\vdash \widehat{B} \to \Box(\widehat{YN} \to \alpha)$. It follows that $\vdash \widehat{B} \to (\neg\Box\alpha \to \neg\Box\widehat{YN})$ and, since $\neg\Box\alpha \in B$, we have $\vdash \widehat{B} \to \neg\Box\widehat{YN}$.

To get a contradiction, we now proceed to prove that we also have $\vdash \widehat{B} \to \Box\widehat{YN}$. The proof can be divided into four parts specified as follows.

1. By Axiom **I1**, $\vdash \widehat{DY} \to \boxbar\Box\widehat{DY}$. Together with the construction of DY and $\Delta^{\phi}_{\boxbar}$, this implies that

$$\{\Box\boxminus\alpha \in Cls(\Sigma) \mid \boxminus\alpha \in \Delta \cap Sub(\Sigma)\} \subseteq \Delta^{\phi}_{\boxbar} \subseteq B \ .$$

Thus, $\vdash \widehat{B} \to \Box\widehat{DY}$.

2. By Axiom **I1** and axiom **5** for $\boxminus$, $\vdash \widehat{DN} \to \boxbar\Box\widehat{DN}$. Together with the construction of DN and $\Delta^{\phi}_{\boxbar}$, this implies that

$$\{\Box\neg\boxminus\alpha \in Cls(\Sigma) \mid \neg\boxminus\alpha \in \Delta \cap Sub(\Sigma)\} \subseteq \Delta^{\phi}_{\boxbar} \subseteq B \ .$$

Thus, $\vdash \widehat{B} \to \Box\widehat{DN}$.

3. By Axiom **I1** and Axiom $\boxplus$**1**, $\vdash \widehat{EY} \to \boxbar\Box\widehat{EY}$. Together with the construction of EY and $\Delta^{\phi}_{\boxbar}$, this implies that

$$\{\Box\boxbar\beta \in Cls(\Sigma) \mid \boxbar\beta \in \Delta \cap Sub(\Sigma)\} \subseteq \Delta^{\phi}_{\boxbar} \subseteq B \ .$$

Hence, $\vdash \widehat{B} \to \Box\widehat{EY}$.

4. By Axiom **I1** and Axiom $\boxplus$**2**, $\vdash \widehat{EN} \to \boxbar\Box\widehat{EN}$. Together with the construction of EN and $\Delta^{\phi}_{\boxbar}$, this implies that

$$\{\Box\neg\boxbar\beta \in Cls(\Sigma) \mid \neg\boxbar\beta \in \Delta \cap Sub(\Sigma)\} \subseteq \Delta^{\phi}_{\boxbar} \subseteq B \ .$$

Hence, $\vdash \widehat{B} \to \Box\widehat{EN}$.

Therefore, $\vdash \widehat{B} \to \Box\widehat{YN}$. This is a contradiction, so $\widehat{B}\wedge\Diamond\widehat{A_0}$ must be consistent.

Now, in order to extend the consistent $\widehat{B} \wedge \Diamond\widehat{A_0}$, suppose as an inductive hypothesis that A_n is defined such that $B \wedge \Diamond A_n$ is consistent ($1 \le n \le m$). Then

$$\vdash \Diamond\widehat{A_n} \leftrightarrow \Diamond((\widehat{A_n} \wedge \alpha_{n+1}) \vee (\widehat{A_n}\wedge \sim \alpha_{n+1}))$$

and thus $\vdash \Diamond\widehat{A_n} \leftrightarrow (\Diamond(\widehat{A_n} \wedge \alpha_{n+1}) \vee \Diamond(\widehat{A_n}\wedge \sim \alpha_{n+1}))$. Therefore, either for $A' = A_n \cup \{\alpha_{n+1}\}$ or for $A' = A_n \cup \{\sim \alpha_{n+1}\}$, we have $B \wedge \Diamond A'$ is consistent. By choosing A_{n+1} to be the consistent expansion, and by letting A be A_m, we have that $\widehat{B}\wedge\Diamond\widehat{A}$ is consistent.

For the other direction, suppose there is an α-atom $A \in W$ such that $\neg\alpha \in A \in f(B)$; then $\widehat{B} \wedge \Diamond\widehat{A}$ is consistent, which implies that $\widehat{B} \wedge \Diamond\neg\alpha$ is consistent. Since $\Box\alpha \in Cls(\phi)$ and B is β-atom over ϕ (and hence maximal consistent in $Cls(\phi)$), we must have $\neg \Box \alpha \in B$. $\Box$

C.2.14. Lemma (Existence Lemma for β-formulas). *In the partial MAM over ϕ, for all $[\alpha]\beta \in Cls(\phi)$, $[\alpha]\beta \notin B$ if and only if there is a β-atom $B' \in \mathsf{Ag}$ such that $\neg\beta \in B'$ and $B \twoheadleftarrow^{\|\alpha\|} B'$.*

Proof:
Assume $[\alpha]\beta \notin B$; we will show that $\{\neg\beta\}$ can be extended to an β-atom $B' \in W$ such that $\widehat{B} \wedge \langle\alpha\rangle\widehat{B'}$ is consistent. We construct an appropriate β-atom B' by forcing choices. Enumerate the β-formulas in $Cls(\phi)$ as $\beta_1, \dots, \beta_m$; define B_0 as $\{\neg\beta\} \cup \Delta^{\phi}_{\boxdot}$.

Let $\lambda = \widehat{\Delta^{\phi}_{\boxdot}}$. We first prove that $\widehat{B} \wedge \langle\alpha\rangle\widehat{B_0}$ is consistent. Suppose not. By an argument similar to that in Lemma C.2.13, we can get $\vdash \widehat{B} \to \neg[\alpha]\lambda$. By rule **N** for $\boxdot$, we have (1) $\vdash \boxdot\widehat{B} \to \boxdot\neg[\alpha]\lambda$. However, $\vdash \boxdot\widehat{B} \to \boxdot\lambda$, since $\Delta^{\phi}_{\boxdot} \subseteq B$. By Axiom **I2**, $\vdash \boxdot\lambda \to \boxdot[\alpha]\lambda$, which implies that (2) $\vdash \boxdot\widehat{B} \to \boxdot[\alpha]\lambda$. But (1) and (2) lead to a contradiction, since $\neg \boxdot \bot \in B$. Therefore, $\widehat{B} \wedge \langle\alpha\rangle\widehat{B_0}$ is consistent.

The induction part to extend $\widehat{B} \wedge \langle\alpha\rangle\widehat{B_0}$ is similar to that in the proof of the previous lemma, and so is this lemma's other direction. $\Box$

Now we are ready to prove the truth lemma.

C.2.15. Lemma (Truth Lemma). *Let $\mathcal{M}_{\Delta}^{\{\|\alpha\| | [\alpha]\top \in Cls(\phi)\}}$ be the partial MAM over ϕ. For all $\varphi \in Cls(\phi)$,*

$$\mathcal{M}_{\Delta}^{\{\|\alpha\| | [\alpha]\top \in Cls(\phi)\}}, (A, B) \vDash \varphi \quad \textit{if and only if} \quad \varphi \in A \cup B$$

Proof:
We proceed by induction on the degree of φ.

- For $\top$, $(A, B) \vDash \top$ and $\top \in A$ and $\top \in B$, so it is trivially satisfied. For atomic propositions p, we have $(A, B) \vDash p$ if and only if $A \in V(p)$ if and only if $p \in A$ if and only if $p \in A \cup B$.

- For the Boolean cases, there are two cases: one for α-formula and one for β-formula. In both, the proof is routine.

- For the four modal operators, the proof uses their existence lemmas. Here, as an example, we only deal with the universal modality on the set of arguments. Assume that $(A, B) \vDash \boxplus\beta$; by its truth definition, for all $B' \in \mathsf{Ag}$, $(A, B') \vDash \beta$. By induction hypothesis, $\beta \in B'$ for all $B' \in \mathsf{Ag}$, which implies that $\beta \in \bigcap \mathsf{Ag} \subseteq \Delta^{\phi}_{\boxplus} = A^{\phi}_{\boxplus}$. (Note that $\boxminus A^{\boxminus} \cap Sub(\Sigma) = \boxminus\Delta^{\boxminus} \cap Sub(\Sigma)$ and $\boxplus A^{\boxplus} \cap Sub(\Sigma) = \boxplus\Delta^{\boxplus} \cap Sub(\Sigma)$ imply $A^{\phi}_{\boxplus} = \Delta^{\phi}_{\boxplus}$ by Axiom **I1**, Axiom $\boxplus$**1** and the constructions of both $Cls(\Sigma)$ and $\Delta^{\phi}_{\boxplus}$, using similar arguments specified in (1)-(4) in Lemma C.2.13). Since $\boxplus\beta \in Cls(\phi)$, we have $\boxplus\beta \in A$. For the other direction, use $\bigcap \mathsf{Ag} \supseteq \Delta^{\phi}_{\boxplus}$.

- For the case of Gfp^{α}, the proof goes as follows. For the first direction, assume that $(A, B) \vDash \mathsf{Gfp}^{\alpha}$. Note that $\mathsf{GFP}^{[\![\alpha]\!]} = \{X \in \mathsf{Ag} \mid (A, X) \vDash \mathsf{Gfp}^{\alpha}\}$. Define $D := \mathsf{GFP}^{[\![\alpha]\!]} \cap \mathsf{Def}^{[\![\alpha]\!]}(B)$. And let δ denote $\bigvee_{X \in D} \widehat{X}$. Define $E := \{X \in \mathsf{Ag} \mid \delta \wedge \langle\alpha\rangle\widehat{X}$ is consistent $\}$. And let ε denote $\bigvee_{X \in E} \widehat{X}$.

  Claim one: $X \in E$ if and only if there is a $Y \in D$ such that $Y \twoheadleftarrow^{[\![\alpha]\!]} X$.

  $$\begin{array}{rcl} X \in E & \Leftrightarrow & \delta \wedge \langle\alpha\rangle\widehat{X} \text{ is consistent} \\ & \Leftrightarrow & \text{there is a } Y \in D \text{ such that } \widehat{Y} \wedge \langle\alpha\rangle\widehat{X} \text{ is consistent} \\ & \Leftrightarrow & \text{there is a } Y \in D \text{ such that } Y \twoheadleftarrow^{\|\alpha\|} X \\ & \Leftrightarrow & \text{there is a } Y \in D \text{ such that } Y \twoheadleftarrow^{[\![\alpha]\!]} X \\ & & \text{(by the induction hypothesis } [\![\alpha]\!] = \|\alpha\|) \end{array}$$

  Claim two: for all $ar \subseteq \mathsf{Ag}$ we have $\vdash \bigvee_{X \in ar} \widehat{X} \to [\alpha]\widehat{\Delta^{\phi}_{\boxplus}}$. Suppose not. Then $\bigvee_{X \in ar} \widehat{X} \wedge \neg[\alpha]\widehat{\Delta^{\phi}_{\boxplus}}$ is consistent, which implies that $\bigvee_{X \in ar} \widehat{X} \wedge \langle\alpha\rangle\neg\widehat{\Delta^{\phi}_{\boxplus}}$ is consistent. So there must be one $\beta \in \Delta^{\phi}_{\boxplus}$ such that $\bigvee_{X \in ar} \widehat{X} \wedge \langle\alpha\rangle\neg\beta$ is consistent. Furthermore, there must be one $X \in ar$ such that $\widehat{X} \wedge \langle\alpha\rangle\neg\beta$ is consistent. By the existence lemma for $[\alpha]$, there must be one $Y \in \mathsf{Ag}$ such that $X \twoheadleftarrow^{\|\alpha\|} Y$ and $\neg\beta \in Y$. However, this is impossible, since $\beta \in \Delta^{\phi}_{\boxplus} \subseteq Y$ by the definition of Ag. Therefore, $\vdash \bigvee_{X \in ar} \widehat{X} \to [\alpha]\widehat{\Delta^{\phi}_{\boxplus}}$.

  Claim three: $\vdash \varepsilon \to \langle\alpha\rangle\delta$. Suppose not. Then $\varepsilon \wedge \neg\langle\alpha\rangle\delta$ is consistent. Thus there must be an $X \in E$ such that $\widehat{X} \wedge \neg\langle\alpha\rangle\delta$ is consistent, which implies that there is no $Y \in D$ such that $\widehat{X} \wedge \langle\alpha\rangle\widehat{Y}$ is consistent. Hence, there is no $Y \in D$ such that $X \twoheadleftarrow^{\|\alpha\|} Y$.

  However, by Claim one and $X \in E$, there is a $Z \in D$ such that $Z \twoheadleftarrow^{[\![\alpha]\!]} X$. Since $Z \in D \subseteq \mathsf{GFP}^{[\![\alpha]\!]}$, for all $X \in \mathsf{Ag}$ such that $Z \twoheadleftarrow^{[\![\alpha]\!]} X$, there is $Y \in \mathsf{GFP}^{[\![\alpha]\!]}$

such that $X \twoheadleftarrow^{[\![\alpha]\!]} Y$. Because $X \in \mathsf{Att}^{[\![\alpha]\!]}(B)$, $Y \in \mathsf{Def}^{[\![\alpha]\!]}(B)$. Hence $Y \in D$ and $X \twoheadleftarrow^{\|\alpha\|} Y$. Contradiction!

Claim four: $\vdash \delta \to [\alpha]\varepsilon$. Suppose not. Then $\delta \wedge \langle\alpha\rangle\neg\varepsilon$ is consistent, which implies that there is $X \in D$ such that $\widehat{X} \wedge \langle\alpha\rangle\neg\varepsilon$ is consistent. By Claim two, $\vdash \widehat{X} \to [\alpha]\widehat{\Delta^{\phi}_{\boxplus}}$. By $\vdash [\alpha]\beta \wedge \langle\alpha\rangle\beta' \to \langle\alpha\rangle(\beta \wedge \beta')$, the consistency of $\widehat{X} \wedge \langle\alpha\rangle\neg\varepsilon$ implies that $\widehat{X} \wedge \langle\alpha\rangle(\widehat{\Delta^{\phi}_{\boxplus}} \wedge \neg\varepsilon)$ is consistent. Thus there must be a $Y \in \mathsf{Ag} \setminus E$ such that $\widehat{X} \wedge \langle\alpha\rangle\widehat{Y}$ is consistent. However, this means that $X \twoheadleftarrow^{\|\alpha\|} Y$. Since $X \in D$, by Claim one, it follows that $Y \in E$, contradictory to the fact that $Y \in \mathsf{Ag} \setminus E$. Therefore, $\vdash \delta \to [\alpha]\varepsilon$.

Claim five: $\vdash \delta \to [\alpha]\langle\alpha\rangle\delta$. By rule **N** for $[\alpha]$ on Claim three, we have $\vdash [\alpha]\varepsilon \to [\alpha]\langle\alpha\rangle\delta$. Together with Claim four, it implies that $\vdash \delta \to [\alpha]\langle\alpha\rangle\delta$.

By rule **R** on Gfp^α we have $\vdash \delta \to \mathsf{Gfp}^\alpha$. Since $B \in \mathsf{Def}^{[\![\alpha]\!]}(B)$ by the definition of Def and $B \in \mathsf{GFP}^{[\![\alpha]\!]}$ by the definition of $\mathsf{GFP}^{[\![\alpha]\!]}$ and the assumption that $(A, B) \vDash \mathsf{Gfp}^\alpha$, we have $B \in D = \mathsf{Def}^{[\![\alpha]\!]}(B) \cap \mathsf{GFP}^{[\![\alpha]\!]}$. Recall that $\delta := \bigvee_{X \in D} \widehat{X}$. By $B \in D$, we have $\vdash \widehat{B} \to \delta$. Together with $\vdash \delta \to \mathsf{Gfp}^\alpha$, it follows that $\vdash \widehat{B} \to \mathsf{Gfp}^\alpha$, which gives us that $\mathsf{Gfp}^\alpha \in B$.

For the other direction, assume $\mathsf{Gfp}^\alpha \in A \cup B$ and take $G := \{X \in \mathsf{Ag} \mid \mathsf{Gfp}^\alpha \in X\}$. We first show that G is a postfix point for $d^{\|\alpha\|}$, with $d^{\|\alpha\|}(G) = \{X \in \mathsf{Ag} \mid \text{ for all } Y \text{ such that } X \twoheadleftarrow^{\|\alpha\|} Y, \text{ there is } Z \in G \text{ such that } Y \twoheadleftarrow^{\|\alpha\|} Z\}$.

Take any $X \in G$. Since $\mathsf{Gfp}^\alpha \in X$, by **Unfold** we get $[\alpha]\langle\alpha\rangle \mathsf{Gfp}^\alpha \in X$. If there is no Y such that $X \twoheadleftarrow^{\|\alpha\|} Y$, then $X \in d^{\|\alpha\|}(G)$. If there is Y such that $X \twoheadleftarrow^{\|\alpha\|} Y$, then take an arbitrary one. Since $[\alpha]\langle\alpha\rangle \mathsf{Gfp}^\alpha \in X$, $\neg\langle\alpha\rangle \mathsf{Gfp}^\alpha \notin Y$. By maximality and $\langle\alpha\rangle \mathsf{Gfp}^\alpha \in Cls(\phi)$ we have $\langle\alpha\rangle \mathsf{Gfp}^\alpha \in Y$. So, by the existence lemma for $[\alpha]$, there must be a set Z such that $Y \twoheadleftarrow^{\|\alpha\|} Z$ and $\mathsf{Gfp}^\alpha \in Z$. Thus, $X \in d^{\|\alpha\|}(G)$, and therefore $G \subseteq d^{\|\alpha\|}(G)$.

By the definition of $\mathsf{GFP}^{\|\alpha\|}$, $G \subseteq \mathsf{GFP}^{\|\alpha\|}$. By the inductive hypothesis, $\mathsf{GFP}.d^{\|\alpha\|} = \mathsf{GFP}^{[\![\alpha]\!]}$. Thus, $G \subseteq \mathsf{GFP}^{[\![\alpha]\!]}$, which means that $B \in \mathsf{GFP}^{[\![\alpha]\!]}$. Therefore, by the truth condition of Gfp^α, $(A, B) \vDash \mathsf{Gfp}^\alpha$.

□

The next step is to prove that the partial MAM over ϕ is indeed a partial MAM.

C.2.16. LEMMA. *For all $B \in \mathsf{Ag}$ in $\mathfrak{M}_{\Delta}^{\{\|\alpha\| \mid [\alpha]\top \in Cls(\phi)\}}$, we have $f(B) \neq \varnothing$.*

Proof:
For all $B \in \mathsf{Ag}$ we have $\Box\bot \notin B$. By the existence lemma for $\Box$, there is a α-atom $A \in W$ such that $\top \in A \in f(B)$. Therefore, for all $B \in \mathsf{Ag}$, $f(B) \neq \varnothing$. □

C.2.17. LEMMA. *For all* $\|\alpha\| \in \{\|\alpha\| \mid [\alpha]\top \in Cls(\phi)\}$ *and all* $B_1, B_2 \in \mathsf{Ag}$, *we have* $B_1 \twoheadleftarrow^{\|\alpha\|} B_2$ *if and only if* $B_1 \twoheadleftarrow^{\overline{\|\alpha\|}} B_2$.

Proof:
This follows directly from axiom **1** and the fact that $\{\|\alpha\| \mid [\alpha]\top \in Cls(\phi)\}$ is closed under complement. □

C.2.18. LEMMA. *For all* $\|\alpha\| \in \{\|\alpha\| \mid [\alpha]\top \in Cls(\phi)\}$ *and all* $B_1, B_2 \in \mathsf{Ag}$, *if* $B_1 \twoheadleftarrow^{\|\alpha\|} B_2$ *then either* $f(B_1) \subseteq \|\alpha\|$ *or else* $f(B_1) \subseteq \overline{\|\alpha\|}$

Proof:
From $B_1 \twoheadleftarrow^{\|\alpha\|} B_2$ it follows that $\widehat{B_1} \wedge \langle\alpha\rangle\widehat{B_2}$ is consistent. Thus, $\widehat{B_1} \wedge \langle\alpha\rangle\top$ is consistent, which implies that $\langle\alpha\rangle\top \in B_1$ and hence $\vdash \widehat{B_1} \to \langle\alpha\rangle\top$. Together with axiom **2a**, $\vdash \langle\alpha\rangle\top \to \Box\alpha \vee \Box\neg\alpha$, it implies that $\vdash \widehat{B_1} \to (\Box\alpha \vee \Box\neg\alpha)$.

Now take any $A \in f(B_1)$, we will show that $A \in \|\alpha\|$ implies $f(B_1) \subseteq \|\alpha\|$. Assume both $A \in f(B_1)$ and $A \in \|\alpha\|$; from $A \in f(B_1)$ it follows that $\widehat{B_1} \wedge \Diamond\widehat{A}$ is consistent, so $\widehat{B_1} \wedge \Diamond\alpha$ is consistent by $\vdash \widehat{A} \to \alpha$. By $\Box\alpha \in Cls(\phi)$ and $\vdash \widehat{B_1} \to (\Box\alpha \vee \Box\neg\alpha)$, we have $\Box\alpha \in B_1$, and therefore, $f(B_1) \subseteq \|\alpha\|$.

It can be shown by similar argument that if $A \notin \|\alpha\|$ then $f(B_1) \subseteq \|\neg\alpha\|$. □

C.2.19. LEMMA. *For all* $\|\alpha\| \in \{\|\alpha\| \mid [\alpha]\top \in Cls(\phi)\}$ *and all* $B_1, B_2 \in \mathsf{Ag}$, *if* $B_1 \twoheadleftarrow^{P} B_2$ *and* $f(B_1) \subseteq \|\alpha\|$, *then* $f(B_2) \subseteq \overline{\|\alpha\|}$.

Proof:
Assume both $B_1 \twoheadleftarrow^{\|\alpha\|} B_2$ and $f(B_1) \subseteq \|\alpha\|$. It follows that $\Box\alpha \in B_1$. Thus, by axiom **2b** we have $\vdash \widehat{B_1} \to [\alpha]\,\Box\,\neg\alpha$ and hence, for all B' such that $B_1 \twoheadleftarrow^{\|\alpha\|} B'$, we have $\nvdash \widehat{B'} \to \neg\,\Box\,\neg\alpha$. Since $\Box\neg\alpha \in Cls(\phi)$, we also have $\Box\neg\alpha \in B_2$. This implies that, for all $A \in f(B_2)$, $\nvdash \widehat{A} \to \alpha$. Thus, for all $A \in f(B_2)$, $\neg\alpha \in A$ and therefore $f(B_2) \subseteq \|\neg\alpha\|$. □

C.2.20. LEMMA. *For all* $P, Q \in \{\|\alpha\| \mid [\alpha]\top \in Cls(\phi)\}$ *and* $B_1, B_2 \in \mathsf{Ag}$, *if* $B_1 \twoheadleftarrow^{\|\alpha\|} B_2$ *and* $f(B_1) \subseteq Q \subseteq P$, *then* $B_1 \twoheadleftarrow^{Q} B_2$.

Proof:
Take $P = \|\alpha'\|$ and $Q = \|\alpha\|$. We claim that for all $A \in W$, $\vdash \widehat{A} \to \boxminus(\alpha \to \alpha')$. Suppose not. Then $\widehat{A} \wedge \diamondminus(\alpha \wedge \neg\alpha')$ is consistent, and thus $\{\alpha, \neg\alpha'\}$ can be extended to an α-atom Y over ϕ belonging to W. However. this contradicts $Q \subseteq P$. Hence for all $A \in W$, $\vdash \widehat{A} \to \boxminus(\alpha \to \alpha')$. This implies that $\vdash \widehat{B_1} \to \Box\,\boxminus\,(\alpha \to \alpha')$.

Now, $B_1 \twoheadleftarrow^{P} B_2$ implies that $\widehat{B_1} \wedge \langle\alpha'\rangle\widehat{B_2}$ is consistent, and $f(B_1) \subseteq Q$ implies that $\vdash \widehat{B_1} \to \Box\alpha$. Together with $\vdash \widehat{B_1} \to \Box\,\boxminus\,(\alpha \to \alpha')$ and Axiom **3**, they imply that $\widehat{B_1} \wedge \langle\alpha\rangle\widehat{B_2}$ is consistent. Therefore, $B_1 \twoheadleftarrow^Q B_2$. $\square$

We have proved that the canonical model over ϕ, $\mathcal{M}_\Delta^{\{\|\alpha\| | [\alpha]\top \in Cls(\phi)\}}$, is indeed a partial MAM. By Lemma C.2.2, this structure can be extended to a multi-agent argumentation model. It is only left to prove that this extension, denoted by $\mathcal{M}_\Delta^F$, indeed preserves the behaviour of the relevant formulas.

C.2.21. Lemma. *For all $\varphi \in Cls(\phi)$,*

$$\mathcal{M}_\Delta^F, (A, B) \vDash \varphi \quad \textit{if and only if} \quad \mathcal{M}_\Delta^{\{\|\alpha\| | [\alpha]\top \in Cls(\phi)\}}, (A, B) \vDash \varphi$$

Proof:
The proof proceeds by induction on the degree of φ. The basic case is trivial since the extension does not change V. The proof for other cases is also routine, as neither the support functions f nor the attack relations $\twoheadleftarrow^{[\![\alpha]\!]}$ for $[\alpha]\top \in Cls(\phi)$ are changed when building the extension. $\square$

Thus,

C.2.22. Theorem. *For all $\varphi \in \mathcal{L}$, $\vDash \varphi$ implies $\vdash \varphi$.*

Appendix D

Appendix of Chapter Five

D.1 Two Equivalent Ways of Defining Apeeriodicity

We first present a result about the equivalence of our definition of aperiodicity in Definition 5.4.9 to another definition of aperiodicity (see the third statement in Corollary D.1.2) in the literature given that the probability matrix is strongly connected.

D.1.1. Lemma. *For each m-by-m probability matrix $\mathbf{T}$, if it is strongly connected, then for all $i, j \in \mathbb{N}$ such that $1 \le i, j \le m$,*

$$g(i) = g(j)$$

where $g(i)$ is the greatest common divisor of the members in the set $\{n \in \mathbb{N} \mid \mathbf{T}^n_{ii} > 0\}$.

Proof:
Assume that the given probability matrix is strongly connected. If $i = j$, it is trivial. So we take $i \neq j$.

The strong connectedness implies that there exists natural numbers l, k such that $\mathbf{T}^l_{ij} > 0$ and $\mathbf{T}^k_{ji} > 0$. This implies that $\mathbf{T}^{l+k}_{ii} \ge \mathbf{T}^l_{ij}\mathbf{T}^k_{ji} > 0$. Hence $g(i)$ divides $l + k$.

Now take any natural number n such that $\mathbf{T}^n_{jj} > 0$. Then $\mathbf{T}^{l+n+k}_{ii} > \mathbf{T}^l_{ij}\mathbf{T}^n_{jj}\mathbf{T}^k_{ji} > 0$. Hence $g(i)$ divides $l+n+k$. Together with $g(i)$ dividing $l+k$, it implies that $g(i)$ divides n. So $g(i)$ is a common divisor of the members in the set $\{n \in \mathbb{N} \mid \mathbf{T}^n_{jj} > 0\}$.

Since $g(j)$ is the greatest common divisor of the members in the set $\{n \in \mathbb{N} \mid \mathbf{T}^n_{jj} > 0\}$,

$$g(i) \le g(j) \quad .$$

Switching the order of i and j in the above argument, we get

$$g(j) \le g(i) \ .$$

Therefore $g(i) = g(j)$. □

The following Corollary follows from the above lemma.

D.1.2. COROLLARY. *Assume that the m-by-m probability matrix is strongly connected, then the following three statements are equivalent:*

1. *for all natural number* $i \le m$, $g(i) = 1$*;*
2. *there is a natural number* $i \le m$ *such that* $g(i) = 1$*;*
3. *the greatest common divisor of the members of the set*

$$\bigcup_{0 \le i \le m} \{n \in \mathbb{N} \mid \mathbf{T}_{ii}^n > 0\}$$

is 1.

Proof:
From 1 to 2, it is trivial. From 2 to 3, it is also obvious. From 3 to 1, suppose 1 does not hold. Take the number i such that $g(i) = x > 1$. Given Lemma D.1.1, for all $j \le m$, $g(j) = x$. So obviously, gcd $\bigcup_{0 \le i \le m} \{n \in \mathbb{N} \mid \mathbf{T}_{ii}^n > 0\} = x > 1$. This completes the proof. □

D.2 An Alternative Proof of the Convergence Result for Regular Markov Chain

In this Appendix, we give an alternative proof of the classical convergence result for the regular Markov chain

We first define regular Markov chain and restate the theorem we are going to prove in this section.

D.2.1. DEFINITION. A Markov chain is called a *regular* chain if some power of its transition matrix has only positive elements.

D.2.2. THEOREM (Theorem 11.7 in Grinstead and Snell (1997)). *Let* $\mathbf{P}$ *be the transition matrix for a regular chain. Then as* $n \to \infty$*, the powers* $\mathbf{P}^n$ *approach a limiting matrix* $\mathbf{W}$ *with all rows the same vector* $\mathbf{w}$*. The vector* $\mathbf{w}$ *is a strictly positive probability vector (i.e. the components are all positive and sum to one).*

Since a trust matrix is a probability matrix, it can also serve as a transition matrix for a Markov chain. We will say a trust matrix a "regular" if it is a transition matrix for a regular Markov chain.

D.2.1 First step of the proof

The first step of the proof is to show that if the trust matrix $\mathbf{T}$ is regular then its generated transition matrix $\mathbb{T}$ is a transition matrix for an absorbing Markov chain,

D.2.3. LEMMA. *Given an indeterministic binary DeGroot model $\mathcal{ID}$, if the trust matrix in the IBDM is regular, then given any vector $\mathbf{b} \in \mathfrak{b}$ with some $i \in G$ such that $\mathbf{b}_i = 1$, there exists a path from $\mathbf{b}$ to $\mathbf{1}$. I.e. there exists a sequence $p^0, p^1, \dots, p^n$ of vectors in $\mathfrak{b}$ such that $p^0 = \mathbf{b}$, $p^n = \mathbf{1}$ and for all natural numbers $k \in [0, n-1]$, $\mathbb{T}_{p^k, p^{k+1}} > 0$.*

Proof:
Given an indeterministic binary DeGroot model $\mathcal{ID} = (G, \mathbf{T}, \mathfrak{b})$, assume that the trust matrix is regular. Take any binary vector $\mathbf{b} \in \mathfrak{b}$ such that $\mathbf{b}_i = 1$ for some $i \in G$. We fix such an agent, ι, satisfying $\mathbf{b}_\iota = 1$. We try to find a natural number n and $n+1$ binary vectors $\mathbf{p}^0, \dots, \mathbf{p}^n$ such that $\mathbb{T}_{p^k, p^{k+1}} > 0$ for all natural numbers $k \in [0, n-1]$ and $p^0 = \mathbf{b}$ and $p^n = \mathbf{1}$.

We first show how to find the sequence of binary vectors. Let $p^0 = \mathbf{b}$. Suppose p^k has been defined,

$$p_i^{k+1} = \begin{cases} 1 & \text{there is } y \in G \text{ such that } \mathbf{T}_{iy} > 0 \text{ and } p_y^k = 1 \\ 0 & \text{otherwise} \end{cases} \tag{D.1}$$

Claim: This definition implies that for all positive integers k, $\mathbb{T}_{p^k, p^{k+1}} > 0$.

Proof: Take any $i \in G$.

If there is $y \in G$ such that $\mathbf{T}_{iy} > 0$ and $p_y^k = 1$, according to the definition of $P(Bv'_i = 1 | Bv = p^k)$ given by equation 5.1, $\mathbf{T}_{iy} > 0$ and $p_y^k = 1$ imply that $P(Bv'_i = 1 | Bv = p^k) > 0$. Since in this case $p_i^{k+1} = 1$ by equation D.1, it follows that $P(Bv'_i = p_i^{k+1} | Bv = p^k) > 0$.

If there is no $y \in G$ such that $\mathbf{T}_{iy} > 0$ and $p_y^k = 1$, then $P(Bv'_i = 1 | Bv = p^k) = 0$, which implies that $P(Bv'_i = 0 | Bv = p^k) = 1 > 0$. By equation D.1, in this case $p_i^{k+1} = 0$. Hence $P(Bv'_i = p_i^{k+1} | Bv = p^k) > 0$.

Therefore, for all $i \in G$, $P(Bv'_i = p_i^{k+1} | Bv = p^k) > 0$, which implies that

$$\mathbb{T}_{p^k, p^{k+1}} = \prod_{i \in G} P(Bv'_i = p_i^{k+1} | Bv = p^k) > 0$$

.

Next, we show that by taking any number n satisfying $\mathbf{T}_{ij}^n > 0$ for all $i, j \in G$ (the existence of such n's are ensured by Definition D.2.1), $p^n = \mathbf{1}$.

Assume that n is the number satisfying $\mathbf{T}_{ij}^n > 0$ for all $i, j \in G$. Then for all $x \in G$, there must be a walk from x to ι in the associated graph $\mathcal{SN}_\mathbf{T}$ whose length is n. For each $x \in G$, fix such a walk w_x from x to ι. Let $S_n = \{w_x \mid x \in G\}$. Keep in mind that for all walks in the associated graph, their length is n and they start from x and end at ι.

Claim: $p^n = \mathbf{1}$. Proof: For all $x \in G$, let w_x^m be the agent m steps away from ι along the path $w_x \in S_n$. For example, $w_x^0 = \iota$.

We first show that for all $x \in G$ and all $m > 0$, $p^m_{w_x^m} = 1$ by induction.

Obviously, $p^1_{w_x^1} = 1$. Because there is $y \in G$, which is w_x^0 (i.e., ι), satisfying $\mathbf{T}_{w_x^1 y} > 0$ and $p_y^0 = 1$, by the definition in equation D.1.

Now suppose that $p^m_{w_x^m} = 1$. By noticing that w_x^m satisfies $\mathbf{T}_{w_x^{m+1} w_x^m} > 0$, we know that $p^{m+1}_{w_x^{m+1}} = 1$ by equation D.1.

Next, we show that for the number n, $G = \{w_x^n \mid x \in G\}$. Recall the definition of w_x. For all $x \in G$, it follows by the definition of w_x that $w_x^n = x$.

Therefore, for all $x \in G$, $p_x^n = p^n_{w_x^n} = 1$. That is $p^n = \mathbf{1}$. This completes the proof. □

The above lemma, together with Definition 5.4.2, Theorem 5.4.3 and Corollary 5.4.7, leads to the the following theorem, which accomplishes the first step of the proof.

D.2.4. THEOREM. *Given an indeterministic binary DeGroot model $\mathcal{ID}$, if the trust matrix in the IBDM is regular, then the transition matrix for the IBDM is a transition matrix for an absorbing Markov chain including only two absorbing states. Hence the powers of the transition matrix converge.*

D.2.2 Second step of the proof

The second step is to prove that the convergence of the powers of the generated transition matrix can ensure the convergence of powers of the trust matrix. This is accomplished in the proof of Corollary 5.5.5, which relies on the proof of Theorem 5.5.3.

Proof of Theorem 5.5.3 Proof:
The proof is by induction. We first prove the base case for all $\mathbf{b}, \mathbf{s} \in \mathfrak{b}$ and all $i \in G$,

$$\mathbf{T}_{i*} \cdot \mathbf{b} = \sum_{\mathbf{s}_i = 1} \mathbb{T}_{\mathbf{bs}} \quad .$$

Claim:

$$\sum_{\{\mathbf{s} \in \mathfrak{b} | \mathbf{s}_i = 1\}} \prod_{\{x \in G | x \neq i\}} P(Bv'_x = \mathbf{s}_x | Bv = \mathbf{b}) = 1$$

Proof:

Recall that Bv is the random variable for the groups' current belief state and Bv' is the random variable for the group's belief state in the next step. Bv_i and Bv'_i are the random variables for the ith agent's current belief state and the belief state in the next step respectively.

Notice that $P(Bv' = \mathbf{s}|Bv = \mathbf{b}) = \mathbb{T}_{\mathbf{bs}}$. Since $\mathbb{T}_{\mathbf{bs}} = \prod_{x\in G} P(Bv'_x = \mathbf{s}_x|Bv = \mathbf{b})$, for all $\mathbf{s} \in \mathfrak{b}$ such that $\mathbf{s}_i = 1$:

$$
\begin{aligned}
P(Bv' = \mathbf{s}|Bv = \mathbf{b}, Bv'_i = 1) &= \frac{P(Bv' = \mathbf{s}|Bv = \mathbf{b})}{P(Bv'_i = 1|Bv = \mathbf{b})} \\
&= \prod_{\{x\in G|x\neq i\}} P(Bv'_x = \mathbf{s}_x|Bv = \mathbf{b}) \ .
\end{aligned}
$$

Obviously, $\sum_{\{\mathbf{s}\in\mathfrak{b}|\mathbf{s}_i=1\}} P(Bv' = \mathbf{s} \mid Bv = \mathbf{b}, Bv'_i = 1) = 1$. This completes the proof of the claim.

Making use of the claim, we can prove the base case:

$$
\begin{aligned}
\mathbf{T}_{i*}\mathbf{b} &= P(Bv'_i = 1|Bv = \mathbf{b}) && \text{(D.2)}\\
&= P(Bv'_i = 1|Bv = \mathbf{b}) \cdot (\sum_{\{\mathbf{s}\in\mathfrak{b}|\mathbf{s}_i=1\}} \prod_{\{x\in G|x\neq i\}} P(Bv'_x = \mathbf{s}_x|Bv = \mathbf{b})) && \text{(D.3)}\\
&= \sum_{\{\mathbf{s}\in\mathfrak{b}|\mathbf{s}_i=1\}} \prod_{x\in G} P(Bv'_i = \mathbf{s}_x|Bv = \mathbf{b}) && \text{(D.4)}\\
&= \sum_{\{\mathbf{s}\in\mathfrak{b}|\mathbf{s}_i=1\}} \mathbb{T}_{\mathbf{bs}} && \text{(D.5)}
\end{aligned}
$$

Next we prove the case

$$
\mathbf{T}^{n+1}_{i*} \cdot \mathbf{b} = \sum_{\mathbf{s}_i=1} \mathbb{T}^{n+1}_{\mathbf{bs}}
$$

by assuming that

$$
\mathbf{T}^{n}_{i*} \cdot \mathbf{b} = \sum_{\mathbf{s}_i=1} \mathbb{T}^{n}_{\mathbf{bs}} \ .
$$

$$\sum_{\mathbf{s}_i=1} \mathbb{T}^{n+1}_{\mathbf{bs}} = \sum_{\mathbf{s}_i=1} \mathbb{T}^{n}_{\mathbf{b}*} \mathbb{T}_{*\mathbf{s}} \tag{D.6}$$

$$= \mathbb{T}^{n}_{\mathbf{b}*} \cdot \sum_{\mathbf{s}_i=1} \mathbb{T}_{*\mathbf{s}} \tag{D.7}$$

$$= \mathbb{T}^{n}_{\mathbf{b}*} \cdot \begin{bmatrix} \sum_{\mathbf{s}_i=1} \mathbb{T}_{\mathbf{1s}} \\ \vdots \\ \sum_{\mathbf{s}_i=1} \mathbb{T}_{\mathbf{0s}} \end{bmatrix} \tag{D.8}$$

$$= \mathbb{T}^{n}_{\mathbf{b}*} \cdot \begin{bmatrix} \mathbf{T}_{i*}\mathbf{1} \\ \vdots \\ \mathbf{T}_{i*}\mathbf{0} \end{bmatrix} \tag{D.9}$$

$$= \sum_{\mathbf{k}\in\mathfrak{b}} (\mathbb{T}^{n}_{\mathbf{bk}} \cdot \mathbf{T}_{i*} \cdot \mathbf{k}) \tag{D.10}$$

$$= \mathbf{T}^{n}_{i*} \cdot \sum_{\mathbf{k}\in\mathfrak{b}} (\mathbb{T}^{n}_{\mathbf{bk}} \cdot \mathbf{k}) \tag{D.11}$$

$$= \mathbf{T}_{i*} \cdot \begin{bmatrix} \sum_{\mathbf{k}_1=1} \mathbb{T}^{n}_{\mathbf{bk}} \\ \vdots \\ \sum_{\mathbf{k}_{|G|}=1} \mathbb{T}^{n}_{\mathbf{bk}} \end{bmatrix} \tag{D.12}$$

$$= \mathbf{T}_{i*} \cdot \begin{bmatrix} \mathbf{T}^{n}_{1*}\mathbf{b} \\ \vdots \\ \mathbf{T}^{n}_{|G|*}\mathbf{b} \end{bmatrix} \tag{D.13}$$

$$= \mathbf{T}_{i*} \cdot (\mathbf{T}^{n}\mathbf{b}) \tag{D.14}$$

$$= (\mathbf{T}_{i*}\mathbf{T}^{n}) \cdot \mathbf{b} \tag{D.15}$$

$$= \mathbf{T}^{n+1}_{i*}\mathbf{b} \tag{D.16}$$

This completes the proof. □

D.2.3 Third step of the proof

Given a transition matrix for a regular chain $\mathbf{P}$, take $\mathbf{P}$ as the trust matrix in an IBDM.

Generate a transition matrix $\mathbb{P}$ from the transition matrix according to Definition 5.3.3. By Theorem 5.4.3 and Theorem proved in the first step, $\mathbb{P}^n$ approach to a limiting matrix. By Corollary 5.5.5, it implies that $\mathbf{P}^n$ approach a limiting matrix.

The remaining task is to prove that the limiting matrix of $\mathbb{P}^n$ is a matrix whose rows are all the same vector $\mathbf{w}$ and the vector $\mathbf{w}$ is strictly positive probability vector (i.e. the components are all positive and sum to one).

Notice that no matter which i is taken, by Corollary 5.5.8,

$$\mathbf{P}^{\infty}_{ij} = \mathbf{P}^{\infty}_{i*}\mathbf{e}_j = \sum_{\mathbf{s}_i=1} \mathbb{P}^{\infty}_{\mathbf{e}_j\mathbf{s}} = \mathbb{P}^{\infty}_{\mathbf{e}_j\mathbf{1}}$$

where $\mathbf{e}_j$ is the vector where the ith entry is 1 and the other entries are all 0s. This is because $\mathbf{1}$ is the only belief vector where $\mathbf{s}_i = 1$ such that $\mathbb{P}^{\infty}_{\mathbf{e}_j\mathbf{s}} > 0$. (Suppose not. Then there will be other absorbing state other than $\mathbf{1}$ and $\mathbf{0}$. Denote this belief state by $\mathbf{t}$. Then $\mathbb{P}_{\mathbf{tt}} = 1$. This implies that if $\mathbf{t}_i = 0$ then the column vector $\mathbf{P}_{*i}$ is the zero vector. This contradicts the fact that $\mathbf{P}$ is regular.) Hence it follows that the column vector of $\mathbf{P}^{\infty}$ is a constant vector.

$\mathbf{P}$ is a probability matrix, $\mathbf{P}^{\infty}$ must also be a probability matrix. We still need to prove that $\mathbf{P}^{\infty}$ is strictly positive. It follows from the fact that $\mathbb{P}^{\infty}_{\mathbf{e}_j\mathbf{1}} > 0$ for all $\mathbf{e}_j$.

This completes the proof.

D.3 Proof of Lemma 5.4.10

The proof of Lemma 5.4.10 can be achieved by Lemma D.2.3 and the following theorem. Although the following theorem is a known result (Golub and Jackson, 2010, Lemma 2 in Appendix A), which follows from Theorem 1 and Theorem 2 in Perkins (1961), we provide a direct proof in this section.

D.3.1. THEOREM. *For all probability matrices $\mathbf{T}$, $\mathbf{T}$ is a transition matrix for a regular Markov chain if and only if it is strongly connected and aperiodic.*

Proof for Theorem D.3.1

D.3.2. LEMMA. *Assume that the m-by-m probability matrix $\mathbf{T}$ is aperiodic, then there exists $N < \infty$ such that*

$$\mathbf{T}^n_{ii} > 0$$

for all natural numbers $i \leq m$ and all natural numbers $n \geq N$.

Proof:
The proof of this lemma itself relies on the following result in the number theory:

> Let $l_1, l_2, \ldots$ be positive integers with their greatest common divisor equals 1. Then there is an integer L such that for all $l \geq L$ there are non-negative integers $\alpha_1, \alpha_2, \ldots$ such that
>
> $$l = \alpha_1 l_1 + \alpha_2 l_2 + \ldots$$

Take any $i \leq m$ and let $\mathrm{cyc}_i = \{n \in \mathbb{N} \mid \mathbf{T}^n_{ii} > 0\}$.

Notice numbers in cyc_i has greatest common divisor 1. Moreover, for all $n_1, n_2, \ldots \in \mathrm{cyc}_i$, given any non-negative integers $\alpha_1, \alpha_2, \ldots$,

$$\alpha_1 n_1 + \alpha_2 n_2 + \ldots \in \mathrm{cyc}_i \ .$$

Apply the result from the number theory quoted above to numbers in cyc_i. Then there is N such that for all $n \geq N$ there are non-negative integers $\alpha_1, \alpha_2, \ldots$ such that

$$n = \alpha_1 n_1 + \alpha_2 n_2 + \ldots$$

where $n_1, n_2, \ldots \in \mathrm{cyc}_i$. Together with our above observation, it follows immediately that $n \in \mathrm{cyc}_i$. This completes the proof.

□

With this lemma, we can prove Theorem D.3.1.

Proof:

We first prove that if the m-by-m probability matrix $\mathbf{T}$ is a transition matrix for a regular Markov chain, then it is strongly connected and aperiodic.

Assume that $\mathbf{T}$ is a transition matrix for a regular Markov chain. It is easy to see why the strong connectedness follows. So we focus on the aperiodicity. Take an arbitrary i satisfying $0 \leq i \leq m$. There is a natural number N such that for all $n \geq N$, $\mathbf{T}^n_{ii} > 0$. The greatest common divisor of any two numbers n and $n+1$ is 1. So it follows immediately that for all i such that $0 \leq i \leq m$, $g(i) = 1$. This completes the proof for one direction.

For the other direction, assume that the m-by-m probability matrix $\mathbf{T}$ is strongly connected and aperiodic.

Take any i, j, denote the shortest path from i to j by s_{ij}. The existence of the path is ensured by the strong connectedness. Let $ls = \max\{s_{ij} \mid 0 \leq i, j \leq m \text{ and } j \neq i\}$ be the largest length of all the shortest paths between two different nodes in the associated graph of $\mathbf{T}$.

By Lemma D.3.2, there is a natural number N such that for all $n \geq N$ and all $0 \leq i \leq m$

$$\mathbf{T}^n_{ii} > 0 \ .$$

Because $\mathbf{T}$ is assumed to be aperiodic.

We claim that for all natural numbers $n \geq ls + N$ and any $0 \leq i, j \leq m$

$$\mathbf{T}^n_{ij} > 0 \ .$$

The proof for this claim is not difficult once the following fact is realized: given any i, in order to reach j by n steps where $n \geq ls + N$, we can first take the shortest path from i to j, and then take the cycle with $N + (ls - s_{ij})$ or more steps to end at node j.

This completes the proof of the other direction and thus the whole proof.

□

D.4 Background on Absorbing Markov Chain

We first sketch the idea of how to prove Theorem 5.4.3, which helps us get a better understanding of the transition matrix for an absorbing Markov chain. Then we introduce the fundamental matrix for the absorbing Markov chain, which encodes a lot of key information about the absorbing Markov chain. All the material presented in this part can be found in most of the textbooks on Markov chain theory. Our presentation is based on Grinstead and Snell (1997).

Take the transition matrix $\mathbb{P}$ for an arbitrary absorbing Markov chain. Assume that there are r absorbing states. The other states are called *transient states* and let t be the number of transient states. The transition matrix $\mathbb{P}$ can always be written in the following form by permuting the rows and columns.

$$\mathbb{P} = \left(\begin{array}{c|c} \mathbf{Q} & \mathbf{R} \\ \hline \mathbf{Z} & \mathbf{I} \end{array}\right) \tag{D.17}$$

where $\mathbf{Q}$ is a t-by-t matrix, $\mathbf{R}$ is a non-zero t-by-r matrix, $\mathbf{Z}$ is a r-by-t zero matrix and $\mathbf{I}$ is a r-by-r identity matrix. It is called the *canonical form* of the transition matrix for an absorbing Markov chain. For example, the following matrix is in the canonical form:

$$\mathbb{P} = \left(\begin{array}{ccc|cc} 0 & 1/2 & 0 & 1/2 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 1/2 \\ \hline 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right) \tag{D.18}$$

A standard matrix algebra argument shows that

$$\mathbb{P}^n = \left(\begin{array}{c|c} \mathbf{Q}^n & * \\ \hline \mathbf{Z} & \mathbf{I}^n \end{array}\right) \tag{D.19}$$

where $*$ stands for the t-by-r matrix in the upper right-hand corner of $\mathbb{P}^n$.

Proving Theorem 5.4.3 is thus equivalent to proving the following theorem.

D.4.1. THEOREM. *Assume that $\mathbb{P}$ is the transition matrix for an absorbing Markov chain and $\mathbf{Q}$ is the upper left matrix in the canonical form of $\mathbb{P}$. Then $\lim_{n\to\infty}\mathbf{Q}^n = \mathbf{Z}$, where $\mathbf{Z}$ is a t-by-t zero matrix.*

We do not prove this theorem here. The point of introducing the canonical form and Theorem D.4.1 is to illustrate how $\mathbb{P}^\infty$ looks. It is quite clear that in $\mathbb{P}^\infty$ the probability is distributed only among the absorbing states given any initial states.

Moreover, by taking $\mathbf{N} = \mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \ldots$, we get the fundamental matrix of the absorbing matrix, whose existence can be proved. Then what does this fundamental matrix tell us? The entry $\mathbf{N}_{ij}$ tells us the expected numbers of times the absorbing Markov process would pass through the transient states j given that the initial state is the transient state i. And if we take the sum $\sum_{j=1}^{t} \mathbf{N}_{ij}$, it tells us the expected number of steps the process would take before finally being absorbed given that the initial state is the transient state i.

In Section 5.5, we show how we can compute the probability $\mathbb{T}^{\infty}_{\mathbf{b1}}$ via the trust matrix $\mathbf{T}$. There is another way of computing the absorption probability. Just multiply the fundamental matrix $\mathbf{N}$ with the upper right matrix $\mathbf{R}$ in the canonical form of the transition matrix. $(\mathbf{NR})_{ij}$ will give us the probability of the process being absorbed by the absorbing state j given that it starts with the transient state i.

At last, it is worth mentioning that it is not difficult to compute the fundamental matrix $\mathbf{N}$. Because it can be proved that

$$\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1} \ .$$

D.5 Proof of Theorem 5.7.7

What we need to prove in Theorem 5.7.7 is that given a set of formulas $\Phi \subseteq \mathcal{L}_{B\succcurlyeq}$ and a formula $\phi \in \mathcal{L}_{B\succcurlyeq}$, ϕ can be derived from Φ in the axiom system KS if and only if for all regular KDMs $\mathcal{M}$ and all possible worlds w in $\mathcal{M}$, if all the formulas in Φ are satisfied on w then ϕ is also satisfied on w.

The proof for the direction from left to right (i.e., the soundness) is routine. We only need to prove that all the axioms in KS are valid and the rules in KS preserves validity. The details are left out here.

We mainly sketch the proof for the direction from right to left (i.e., the strong completeness) in this appendix. We follow the standard strategy of proving strong completeness (see Blackburn, Rijke and Venema (2002, Chapter 4)). That is, proving that given any consistent set of formulas $\Sigma \subseteq \mathcal{L}_{B\succcurlyeq}$, there is a regular KDM such that Σ can be satisfied on a certain possible world in it. So the core of the proof is to construct a model meeting all the requirements.

To construct such a model, we need the possible worlds. Following the standard strategy, we first collect all the maximal consistent set (MCS) and prove the Lindenbaum's Lemma. That is, if Σ is consistent then there is a MCS Σ^+ such that $\Sigma \subseteq \Sigma^+$. We leave the proof for this lemma out here and refer readers to Blackburn, Rijke and Venema (2002, Lemma 4.17).

With the collection of all the MCSs (denoted by MCS) and the existence of Σ^+, we define the model, denoted by M^{Σ}. Let $\Gamma_{\succcurlyeq}$ be the set $\{C \succcurlyeq D \mid C \succcurlyeq D \in \Gamma\}$ for $\Gamma \in$ MCS. The definition for W^{Σ} and the definition for R_i^{Σ} are as follows:

- $W^{\Sigma} = \{\Gamma \in \text{MCS} \mid \Sigma^+_{\succcurlyeq} = \Gamma_{\succcurlyeq}\}$;
- $R_i^{\Sigma}\Delta\Gamma$ if and only if for all formulas $\varphi \in \mathcal{L}_{B\succcurlyeq}$, $B_i\varphi \in \Delta$ implies $\varphi \in \Gamma$.

Next, we construct the trust matrix. This is the part not included in the standard proof, so it needs extra labour. First, we apply the following theorem from Scott (1964) to prove the existence of a strictly positive probability measure on the power set of G satisfying all the formulas of the form $C \succcurlyeq D$ in Σ where $C, D \subseteq G$.

D.5.1. Theorem (Theorem 4.1 in Scott (1964)). *Let Bl be a finite Boolean algebra and let $\succcurlyeq$ be a binary relation on Bl. For $\succcurlyeq$ to be realizable by a probability measure on Bl it is necessary and sufficient that the conditions*

1. $\top \succ \bot$,
2. $x \succcurlyeq \bot$,
3. $x \succcurlyeq y$ *or* $y \succcurlyeq x$,
4. $x_0 + x_1 + \cdots + x_{n-1} = y_0 + y_1 + \cdots + y_{n-1}$ *implies* $y_0 \succcurlyeq x_0$

hold for all $x, y \in Bl$ and all sequences $x_0, \ldots, x_{n-1}, y_0, \ldots, y_{n-1} \in Bl$ where $x_i \succcurlyeq y_i$ for $i < n, i > 0$, and $n > 0$.

The equation $x_0 + x_1 + \cdots + x_{n-1} = y_0 + y_1 + \cdots + y_{n-1}$ in the fourth condition means exactly what is described in the antecedent of the axiom **Scott**. Review axioms **SP**, **CO** and **Scott** in Table 5.1. By Theorem D.5.1, it is easy to see that they are sufficient to ensure the existence of a strictly positive probability measure on the power set of G which realizes $\succcurlyeq$.

Take a strictly positive probability measure on the power set of G which realises $\succcurlyeq$. Let it be the row vector for each row of the trust matrix. Hence we get a trust matrix for a regular community. The trust matrix $\mathbf{T}^\Sigma$ we construct is very special in the sense of satisfying

$$(\mathbf{T}^\Sigma)^\infty = \mathbf{T}^\Sigma \ .$$

Since $\mathbf{T}^\Sigma$ is the trust matrix for a regular community, we can conclude that the model M^Σ in the following definition is indeed a regular Kripke-DeGroot model.

D.5.2. Definition. The model M^Σ for a set of consistent formulas σ is the structure $(W^\Sigma, R_i^\Sigma, \mathbf{T}^\Sigma, V^\Sigma)$ where W^Σ, R_i^Σ and $\mathbf{T}^\Sigma$ are defined as we described above and

$$V^\Sigma(p) = \{\Gamma \in W^\Sigma \mid p \in \Gamma\} \ .$$

We have constructed the regular Kripke-DeGroot model, the remaining work is to prove the truth lemma: for all $\varphi \in \mathcal{L}_{B\succcurlyeq}$ and all $\Gamma \in W^\Sigma$, $\varphi \in \Gamma$ if and only if $M^\Sigma, \Gamma \vDash \varphi$.

D.5.3. Lemma (Truth Lemma). *For all formula $\varphi \in \mathcal{L}_{B\succcurlyeq}$ and all $\Gamma \in W^\Sigma$,*

$$M^\Sigma, \Gamma \vDash \varphi \textit{ iff } \varphi \in \Gamma \ .$$

Proof:
The proof goes by induction on the formula φ.

The case where φ is an atomic proposition is trivial by the definition of V. The case where φ is of the form $\neg\psi$ or $\psi \wedge \chi$ is also straightforward. The proof for the case where φ is of the form $B_i\psi$ is standard as in Blackburn, Rijke and Venema (2002, Lemma 4.21), with a little modification to take care of the fact that in our model W^Σ is not the whole set MCS but rather $\{\Gamma \in \text{MCS} \mid \Sigma^+_{\succcurlyeq} = \Gamma_{\succcurlyeq}\}$;

The only thing needs extra attention for this case here is that the set of possible worlds W^Σ is not the set of all the MCSs.

The case where φ is of the form $C \succcurlyeq D$ is proved by invoking the fact that $\mathbf{T}^\Sigma$ realises all the formulas of the form $C \succcurlyeq D$.

$\square$

Because there is a possible world Σ^+ in W^Σ such that Σ is a subset, we thus have the desired result: for all $\varphi \in \Sigma$, $M^\Sigma, \Sigma^+ \vDash \varphi$.

D.5.4. Theorem. *Given any consistent set of formulas Σ in $\mathcal{L}_{B\succcurlyeq}$, there is a regular Kripke-DeGroot model such that Σ is satisfiable in it.*

This completes the proof for the strong completeness.

Bibliography

Arnold, André and Damian Niwiński. 2001. *Rudiments of μ-Calculus*. 1st edition ed. Amsterdam: North Holland.

Baltag, Alexandru and Andrés Occhipinti. 2017. Evidence logics with relational evidence. In *Logic, Rationality, and Interaction: 6th International Workshop, LORI 2017, Sapporo, Japan, September 11-14, 2017, Proceedings*, ed. Alexandru Baltag, Jeremy Seligman and Tomoyuki Yamada. Berlin, Heidelberg: Springer pp. 17–32.

Baltag, Alexandru, Bryan Renne and Sonja Smets. 2014. "The logic of justified belief, explicit knowledge, and conclusive evidence." *Annals of Pure and Applied Logic* 165(1):49–81.

Baltag, Alexandru, Nick Bezhanishvili, Aybüke Özgün and Sonja Smets. 2016*a*. Justified belief and the topology of evidence. In *Logic, Language, Information, and Computation*, ed. Jouko Väänänen, Åsa Hirvonen and Ruy de Queiroz. Springer, Berlin, Heidelberg pp. 83–103.

Baltag, Alexandru, Nick Bezhanishvili, Aybüke Özgün and Sonja Smets. 2016*b*. Justified Belief and the Topology of Evidence. Technical report Institute for Logic, Language and Computation, University of Amsterdam.

Baltag, Alexandru and Sonja Smets. 2008. A qualitative theory of dynamic interactive belief revision. In *Logic and the foundations of game and decision theory (LOFT 7)*, ed. Wiebe van der Hoek Giacomo Bonanno and Michael Wooldridge. Vol. 3 of *Text in Logic and Games* Amsterdam University Press pp. 9–58.

Baltag, Alexandru, Virginie Fiutek and Sonja Smets. 2016. Beliefs and evidence in justification models. In *Advances in Modal Logic 11, proceedings of the 11th conference on "Advances in Modal Logic" held in Budapest, Hungary, August 30 - September 2, 2016.* pp. 156–176.

Baltag, Alexandru, Zoé Christoff, Rasmus K Rendsvig and Sonja Smets. 2016. Dynamic epistemic logics of diffusion and prediction in social networks. In *Proceedings of the 12th Conference on Logic and the Foundations of Game and Decision Theory.*

Baroni, Pietro, Massimiliano Giacomin and Beishui Liao. 2014. "On topology-related properties of abstract argumentation semantics. A correction and extension to Dynamics of argumentation systems: A division-based method." *Artificial Intelligence* 212:104–115.

Berto, Francesco. 2007. *How to Sell a Contradiction.* College Publications.

Blackburn, Patrick, Maarten de Rijke and Yde Venema. 2002. *Modal Logic.* Cambridge: Cambridge University Press.

Boutilier, Craig. 1994*a*. "Conditional logics of normality: a modal approach." *Artificial Intelligence* 68(1):87–154.

Boutilier, Craig. 1994*b*. "Unifying default reasoning and belief revision in a modal framework." *Artificial Intelligence* 68(1):33–85.

Bratman, Michael E. 2014. *Shared Agency: A Planning Theory of Acting Together.* 1st edition ed. New York, NY: Oxford University Press.

Burgess, John P. 1981. "Quick completeness proofs for some logics of conditionals." *Notre Dame Journal of Formal Logic* 22(1):76–84.

Caminada, Martin W. A. and Dov M. Gabbay. 2009. "A logical account of formal argumentation." *Studia Logica* 93(2):109.

Cariani, Fabrizio. 2016. "Local supermajorities." *Erkenntnis* 81(2):391–406.

Cathcart, Thomas and Daniel Klein. 2007. *Plato and a Platypus Walk Into a Bar: Understanding Philosophy Through Jokes.* New York: Abrams Image.

Chellas, Brian F. 1980. *Modal Logic: An Introduction.* Cambridge University Press.

Christoff, Zoé. 2016. Dynamic Logics of Networks – Information Flow and the Spread of Opinion. PhD thesis University of Amsterdam.

Conee, Earl and Richard Feldman. 2004. *Evidentialism.* Oxford: Oxford University Press.

DeGroot, Morris H. 1974. "Reaching a consensus." *Journal of the American Statistical Association* 69(345):118–121.

Dietrich, Franz and Christian List. 2007. "Judgment aggregation by quota rules: majority voting generalized." *Journal of Theoretical Politics* 19(4):391–424.

Dietrich, Franz and Christian List. 2014. From degrees of belief to beliefs: Lessons from judgment-aggregation theory. MPRA Paper University Library of Munich, Germany.

Dokow, Elad and Ron Holzman. 2010. "Aggregation of binary evaluations with abstentions." *Journal of Economic Theory* 145(2):544–561.

Dretske, Fred. 1971. "Conclusive Reasons." *Australasian Journal of Philosophy* 49(1):1–22.

Dryzek, John S. 2000. *Deliberative Democracy and Beyond: Liberals, Critics, Contestations*. Oxford political theory Oxford University Press.

Dung, Phan Minh. 1995. "On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games." *Artificial Intelliegence* 77(2):321–357.

Egré, Paul. 2017. "Knowledge as *de re* true belief?" *Synthese* 194(5):1517–1529.

Égré, Paul, Paul Marty and Bryan Renne. 2014. "Knowledge, justification, and reason-based belief." *CoRR* abs/1412.1862.
URL: *http://arxiv.org/abs/1412.1862*

Fagin, Ronald, Joseph Halpern, Yorav Moses and Moshe Vardi. 1995. *Reasoning About Knowledge*. MIT press.

Foley, Richard. 1992. "The epistemology of belief and the epistemology of degrees of belief." *American Philosophical Quarterly* 29(2):111–124.

Foley, Richard. 2009. Beliefs, degrees of velief, and the Lockean thesis. In *Degrees of Belief*, ed. Franz Huber and Christoph Schmidt-Petri. Number 342 *in* "Synthese Library" Springer Netherlands pp. 37–47.

Gabbay, Dov M., Massimiliano Giacomin, Beishui Liao and Leendert van der Torre. 2016. "Present and future of formal argumentation." *Dagstuhl Reports* 5(8):74–89.

Geffner, Hector. 1992. *Default Reasoning: Causal and Conditional Theories*. Cambridge, MA, USA: MIT Press.

Gettier, Edmund L. 1963. "Is justified true belief knowledge." *Analysis* 23(6):121–123.

Gilbert, Margaret. 1987. "Modelling collective belief." *Synthese* 73(1):185–204.

Goldman, Alvin I. 1967. "A cuasal theory of knowing." *The Journal of Philosophy* 64(12):357–372.

Goles Ch, Eric. 1985. "Dynamics of positive automata networks." *Theoretical Computer Science* 41:19–32.

Golub, Benjamin and Matthew O. Jackson. 2010. "Naïve learning in social networks and the wisdom of crowds." *American Economic Journal: Microeconomics* 2(1):112–149.

Grandi, Umberto, Emiliano Lorini and Laurent Perrussel. 2015. Propositional opinion diffusion. In *Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems*. AAMAS '15 Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems pp. 989–997.

Grinstead, Charles M. and J. Laurie Snell. 1997. *Introduction to Probability*. 2nd revised edition ed. Providence, RI: American Mathematical Society.

Grossi, Davide. 2010*a*. Argumentation in the view of modal logic. In *Argumentation in Multi-Agent Systems - 7th International Workshop, ArgMAS 2010 Revised, Selected and Invited Papers*. pp. 190–208.

Grossi, Davide. 2010*b*. On the logic of argumentation theory. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: Volume 1 - Volume 1*. AAMAS '10 Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems pp. 409–416.

Grossi, Davide. 2012. Fixpoints and iterated updates in abstract argumentation. In *Proceedings of the Thirteenth International Conference on Principles of Knowledge Representation and Reasoning*. KR'12 AAAI Press pp. 65–74.

Grossi, Davide. 2013. "Abstract argument games via modal logic." *Synthese* 190:5–29.

Grossi, Davide and Wiebe van der Hoek. 2014. Justified beliefs by justified arguments. In *Proceedings of the Fourteenth International Conference on Principles of Knowledge Representation and Reasoning*. KR'14 AAAI Press pp. 131–140.

Gutmann, Amy and Dennis Thompson. 2004. *Why Deliberative Democracy?* Princeton University Press.

Gärdenfors, Peter. 1975. "Qualitative probability as an intensional logic." *Journal of Philosophical Logic* 4(2):171–185.

Habermas, Jürgen. 1996. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Studies in contemporary German social thought MIT Press.

Hintikka, Jaakko. 1962. *Knowledge and Belief*. Ithaca: Cornell University Press.

Holliday, Wesley H and Thomas F Icard III. 2013. "Measure semantics and qualitative semantics for epistemic modals." *Semantics and Linguistic Theory* 23:514–534.

Holliday, Wesley Halcrow. 2012. Knowing What Follows: Epistemic Closure and Epistemic Logic. PhD thesis Stanford university.

Jackson, Matthew O. 2010. *Social and Economic Networks*. Princeton, NJ: Princeton University Press.

Jeffrey, Richard C. 1965. *The Logic of Decision*. New York: McGraw-Hill.

Kautz, Henry, Bart Selman and Mehul Shah. 1997. "Referral web: combining cocial networks and collaborative filtering." *Commun. ACM* 40(3):63–65.

Keynes, John Maynard. 1923. *A Tract on Monetary Reform*. Macmillan and Co., limited.

Knaster, Bronisław. 1928. "Un théorème sur les fonctions d'ensembles." *Annales de la Société Polonaise de Mathématiques* 6:133–134.

Kozen, Dexter. 1983. "Results on the propositional mu-calculus." *Theoretical Computer Science* 27:333–354.

Kraft, Charles H., John W. Pratt and A. Seidenberg. 1959. "Intuitive probability on finite sets." *The Annals of Mathematical Statistics* 30(2):408–419.

Kraus, Sarit, Daniel Lehmann and Menachem Magidor. 1990. "Nonmonotonic reasoning, rreferential models and cumulative logics." *Artificial Intelligence* 44(1-2):167–207.

Kurucz, Agi. 2007. Combining modal logic. In *Handbook of Modal Logic*, ed. Patrick Blackburn, Johan F.A.K. van Benthem and Frank Wolter. Vol. 3 Elsevier chapter 15, pp. 869–924.

Kyburg, Henry Elt. 1970. *Conjunctivitis*. Dordrecht: Springer Netherlands pp. 55–82.

Kyburg, Henry Ely. 1961. *Probability and the Logic of Rational Belief*. 1st edition ed. Wesleyan University Press.

Lehrer, Keith and Thomas Paxson. 1969. "Knowledge: undefeated justified true belief." *The Journal of Philosophy* 66(8):225–237.

Leitgeb, Hannes. 2014. "The stability theory of belief." *Philosophical Review* 123(2):131–171.

Leitgeb, Hannes. 2017. *The Stability of Belief: How Rational Belief Coheres with Probability*. Oxford University Press.

Levi, Isaac. 1973. *Gambling with Truth: An Essay on Induction and the Aims of Science*. MIT Press.

Lewis, David. 1973. *Counterfactuals*. Blackwell.

Liao, Beishui. 2013. "Toward incremental computation of argumentation semantics: A decomposition-based approach." *Annals of Mathematics and Artificial Intelligence* 67(3):319–358.

Liao, Beishui, Li Jin and Robert C. Koons. 2011. "Dynamics of argumentation systems: A division-based method." *Artificial Intelligence* 175(11):1790–1814.

Lin, Hanti and Kevin T. Kelly. 2012*a*. "A geo-logical solution to the lottery paradox, with applications to conditional logic." *Synthese* 186(2):531–575.

Lin, Hanti and Kevin T. Kelly. 2012*b*. "Propositional reasoning that tracks probabilistic reasoning." *Journal of Philosophical Logic* 41(6):957–981.

List, Christian. 2012. "The theory of judgment aggregation: an introductory review." *Synthese* 187(1):179–207.

List, Christian and Philip Pettit. 2002. "Aggregating sets of judgments: an impossibility result." *Economics and Philosophy* 18(1):89–110.

List, Christian and Philip Pettit. 2011. *Group Agency: The Possibility, Design, and Status of Corporate Agents*. 1st edition ed. Oxford: Oxford University Press.

Liu, Fenrong. 2008. Changing for the Better: Preference Dynamics and Agent Diversity. PhD thesis University of Amsterdam.

Liu, Fenrong and Emiliano Lorini. 2016. Reasons to believe in a social environment. In *Deontic Logic and Normative Systems: 13th International Conference, DEON 2016*. London: College Publications pp. 155–170.

Liu, Fenrong, Jeremy Seligman and Patrick Girard. 2014. "Logical dynamics of belief change in the community." *Synthese* 191(11):2403–2431.

Milgram, Stanley. 1967. "The small world problem." *Psychology Today* 1(1):61–67.

Nozick, Robert. 1981. *Philosophical Explanations.* Harvard University Press.

Özgün, Aybüke. 2017. Evidence in Epistemic Logic. PhD thesis University of Amsterdam.

Perkins, Peter. 1961. "A theorem on regular matrices." *Pacific J. Math.* 11(4):1529–1533.

Pettit, Philip. 2001. "Deliberative democracy and the discursive dilemma." *Philosophical Issues* 11(1):268–299.

Poljak, Svatopluk and Miroslav Sůra. 1983. "On periodical behaviour in societies with symmetric influences." *Combinatorica* 3(1):119–121.

Ramsey, Frank P. 1926. Truth and probability. In *Ramsey, 1931, The Foundations of Mathematics and other Logical Essays*, ed. R.B. Braithwaite. London: Kegan, Paul, Trench, Trubner & Co., New York: Harcourt, Brace and Company chapter seven, pp. 156–198.

Savage, Leonard J. 1954. *The Foundations of Statistics.* New York: Wiley.

Scott, Dana. 1964. "Measurement structures and linear inequalities." *Journal of Mathematical Psychology* 1(2):233–247.

Segerberg, Krister. 1971. Qualitative probability in a Mmodal setting. In *Studies in Logic and the Foundations of Mathematics*, ed. J. E. Fenstad. Vol. 63 of *Proceedings of the Second Scandinavian Logic Symposium* Elsevier pp. 341–352.

Seligman, Jeremy, Fenrong Liu and Patrick Girard. 2013. Facebook and epistemic logic of friendship,. In *Theoretical Aspects of Rationality and Knowledge Proceedings of the 14th Conference*, ed. Burkhard C. Schipper. pp. 229–238.

Shi, Chenwei, Sonja Smets and Fernando R. Velázquez-Quesada. 2017*a*. "Argument-based belief in topological structures." *Electronic Proceedings in Theoretical Computer Science* 251:489–503. In Proceedings TARK 2017, arXiv: 1707.08762.

Shi, Chenwei, Sonja Smets and Fernando R. Velázquez-Quesada. 2017*b*. "Beliefs supported by binary arguments." To appear in Journal of Applied Non-classical Logics.

Shoham, Yoav. 1987. A semantical approach to nonmonotonic logics. In *Readings in Nonmonotonic Reasoning*, ed. Matthew L. Ginsberg. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. pp. 227–250.

Stalnaker, Robert C. 1968. "A theory of conditionals." *Americal Philosophical Quarterly* pp. 98–112.

Tarski, Alfred. 1955. "A lattice-theoretical theorem and its applications." *Pacific Journal of Mathematics* 5(2):285–309.

Toulmin, Stephen. 1958. *The Uses of Argument.* Cambridge University Press.

Travers, Jeffrey and Stanley Milgram. 1969. "An experimental study of the small world problem." *Sociometry* 32(4):425–443.

van Benthem, Johan. 2007. "Dynamic logic for belief revision." *Journal of Applied Non-Classical Logics* 17(2):129–155.

van Benthem, Johan. 2011. *Logical Dynamics of Information Flow.* Cambridge University Press.

van Benthem, Johan. 2015*a*. In between high and low rationality. In *Proceedings of the 3rd International Workshop on Strategic Reasoning*, ed. J. Gutierrez, F. Mogavero, A. Murano and M. Wooldridge.

van Benthem, Johan. 2015*b*. "Oscillations, logic, and dynamical systems." *The Facts Matter. Essays on Logic and Cognition in Honour of Rineke Verbrugge* pp. 9–22.

van Benthem, Johan and Darko Sarenac. 2004. The geometry of knowledge. Technical report ILLC. PP-2004-20.

van Benthem, Johan, David Fernández-Duque and Eric Pacuit. 2012. "Evidence logic: a new look at neighborhood structures." *Advances in Modal Logic* 9:97–118.

van Benthem, Johan, David Fernández-Duque and Eric Pacuit. 2014. "Evidence and plausibility in neighborhood structures." *Annals of Pure and Applied Logic* 165(1):106–133.

van Benthem, Johan and Eric Pacuit. 2011. "Dynamic logics of evidence-based beliefs." *Studia Logica: An International Journal for Symbolic Logic* 99(1/3):61–92.

van Benthem, Johan and Fenrong Liu. 2007. "Dynamic logic of preference upgrade." *Journal of Applied Non-Classical Logics* 17(2):157–182.

van Benthem, Johan and Ştefan Minică. 2012. "Toward a dynamic logic of questions." *Journal of Philosophical Logic* 41(4):633–669.

Van Ditmarsch, Hans, Wiebe van Der Hoek and Barteld Kooi. 2008. *Dynamic Epistemic Logic.* Vol. 337 Springer Science & Business Media.

van Eijck, Jan and Kai Li. 2017. Conditional belief, knowledge and probability. In *Proceedings Sixteenth Conference on Theoretical Aspects of Rationality and Knowledge, TARK 2017, Liverpool, UK, 24-26 July 2017.* pp. 188–206.

Veltman, Frank. 1985. Logics of Conditionals. PhD thesis University of Amsterdam.

Veltman, Frank. 1996. "Defaults in update semantics." *Journal of Philosophical Logic* 25(3):221–261.

Venema, Yde. 2008. *Lectures on the modal μ calculus.* Institute for Logic, Language and Computation.

Watts, Duncan J. and Steven H. Strogatz. 1998. "Collective dynamics of 'small-world' networks." *Nature* 393(6684):440–442.

Wikipedia. 2017. "Blind men and an elephant — Wikipedia, The Free Encyclopedia.".
URL: *https://en.wikipedia.org/wiki/Blind men and an elephant*

Wong, Eva (translator). 2001. *Lieh-tzu: A Taoist Guide to Practical Living.* 1st edition ed. Boston, Mass.; London: Shambhala.

Xianyi, Yang and Dai Naidie. 2016. *Ancient Chinese Fables(Chinese-English).* Beijing: Foreign language press.

Samenvatting

De Grondslag van Opinievorming

We worden constant geconfronteerd met nieuwe informatie. De informatie waarmee we in aanraking komen en waar onze overtuigingen op gebaseerd zijn is vaak chaotisch en soms zelfs tegenstrijdig. Echter worden onze overtuigingen wel geacht consistent te zijn, daar streven we ten minste naar. Is het mogelijk om consistentie te bereiken en te behouden in onze overtuigingen? Zo ja, hoe extraheren we dan consistente overtuigingen uit inconsistente informatie? Deze vragen proberen we te beantwoorden vanuit het perspectief van een logicus, terwijl we gebruik maken van ideeën en methodes uit andere vakgebieden zoals, bijvoorbeeld, topologie, formele argumentatie theorie, niet-monotone redeneringen en Markov-keten theorie.

Deze dissertatie is gestructureerd rondom twee thema's – de overtuigingen van een enkele handelende persoon en die van een groep. Voor zowel de enkele handelende persoon als de groep nemen we aan dat de vorming van overtuigingen een proces is met het oplossen van inconsistentie als basis. Het verschil tussen de beide situaties is dat in het geval van de enkele handelende persoon bewijsmateriaal als basis wordt genomen voor haar overtuigingen en in het geval van de groep worden de overtuigingen van ieder individueel lid als basis genomen voor de groepsovertuigingen.

In het geval van een enkele handelende persoon analyseren we het oplossen van tegenstrijdig bewijsmateriaal aan de hand van het redeneren van de persoon. Twee vormen van redeneren worden onderzocht – argumentatief en default redeneren. De vraag is dus hoe de handelende persoon haar argumentatief en default redeneren coördineert om volledige consistentie te bereiken in haar overtuigingen. We gebruiken topologische semantiek om bewijsmateriaal te modelleren, formele argumentatie theorie en niet-monotone redeneringen om het probleem aan te pakken en bestuderen de logica van de verkregen noties van 'overtuiging' and hun relatie met betrekking tot bewijsmateriaal.

In het geval van groepsovertuigingen bestuderen we twee manieren om de tegenstrijdige overtuigingen van groepsleden op te lossen. Één is gebaseerd op argumentatie en de ander op opinieverspreiding door sociale invloed. We modelleren deze twee vormen van 'groepsredeneren' door Kripke-semantiek te combineren met, respectievelijk, formele argumentatie theorie en Markov-keten theorie. Dit maakt de weg vrij voor onze logische analyse van de resulterende noties van 'groepsovertuiging'.

Het terugkerende thema van de gehele dissertatie is een spanning tussen het vergaren van meer overtuigingen en overtuiging consistenter laten zijn. We tonen aan dat in het geval van de enkele handelende persoon het mogelijk is om op een zinvolle manier een balans te vinden tussen deze twee. Voor groepsovertuigingen moeten we, naast de hoeveelheid en consistentie, ook beschouwen of de samenvoeging van de overtuigingen van de groepsleden democratisch genoeg verloopt. Dus, de spanning is tussen het vergaren van meer overtuigingen, overtuigingen consistenter laten zijn en de samenvoeging democratischer laten verlopen. Onze modellen laten mogelijke manieren zien om deze spanning te verlichten.

Abstract

Reason to Believe

We are confronted with new information all the time. The information we face and on which our beliefs are based, is often chaotic, disordered, and even contradictory. Yet at the same time, our belief is expected to be consistent. At least, we strive for consistent beliefs. Is it possible for us to achieve and maintain consistency in our beliefs? If yes, how do we manage to extract consistent belief from inconsistent information? We attempt to answer these questions from a logician's perspective, borrowing ideas and techniques from other fields, for example, topology, formal argumentation theory, non-monotonic reasoning and Markov chain theory.

The dissertation is structured around two topics – single-agent belief and group belief. For both a single agent and a group of agents, we take belief formation as a process of resolving the inconsistency in its basis. The difference is that in the case of a single agent, evidence is taken as the basis of her belief and in the case of a group of agents, each group member's belief is taken as the basis of the group's belief.

For a single agent, we understand the process of resolving conflicts between different pieces of evidence to be the agent's reasoning. Two forms of reasoning are investigated – argumentational reasoning and default reasoning. The problem thus becomes how the agent coordinates her default reasoning and argumentational reasoning to achieve full consistency in her beliefs. We employ the topological semantics for evidence, formal argumentation theory and non-monotonic reasoning to tackle the problem and study the logic of the resulting notions of belief and their relationship with evidence.

For group belief, we investigate two ways of resolving conflicts between the different group members' beliefs. One is based on argumentation, and the other is opinion diffusion by social influence. We model these two forms of "group reasoning" by combining the Kripke semantics with formal argumentation theory and Markov chain theory respectively, which paves the way for our logical analysis of the notions of group belief based on them.

Throughout the whole dissertation, the recurrent theme is a tension between believing more and believing more consistently. We demonstrate that in the case of a single agent, it is possible to strike a balance in a meaningful way. For group belief, nonetheless, besides its amount of content and consistency, we also need to consider whether the aggregation of the group members' beliefs is democratic enough. So, the tension is between believing more, believing more consistently, and believing more democratically. Our proposals exemplify possible ways of relieving this tension.

Titles in the ILLC Dissertation Series:

ILLC DS-2009-01: **Jakub Szymanik**
Quantifiers in TIME and SPACE. Computational Complexity of Generalized Quantifiers in Natural Language

ILLC DS-2009-02: **Hartmut Fitz**
Neural Syntax

ILLC DS-2009-03: **Brian Thomas Semmes**
A Game for the Borel Functions

ILLC DS-2009-04: **Sara L. Uckelman**
Modalities in Medieval Logic

ILLC DS-2009-05: **Andreas Witzel**
Knowledge and Games: Theory and Implementation

ILLC DS-2009-06: **Chantal Bax**
Subjectivity after Wittgenstein. Wittgenstein's embodied and embedded subject and the debate about the death of man.

ILLC DS-2009-07: **Kata Balogh**
Theme with Variations. A Context-based Analysis of Focus

ILLC DS-2009-08: **Tomohiro Hoshi**
Epistemic Dynamics and Protocol Information

ILLC DS-2009-09: **Olivia Ladinig**
Temporal expectations and their violations

ILLC DS-2009-10: **Tikitu de Jager**
"Now that you mention it, I wonder...": Awareness, Attention, Assumption

ILLC DS-2009-11: **Michael Franke**
Signal to Act: Game Theory in Pragmatics

ILLC DS-2009-12: **Joel Uckelman**
More Than the Sum of Its Parts: Compact Preference Representation Over Combinatorial Domains

ILLC DS-2009-13: **Stefan Bold**
Cardinals as Ultrapowers. A Canonical Measure Analysis under the Axiom of Determinacy.

ILLC DS-2010-01: **Reut Tsarfaty**
Relational-Realizational Parsing

ILLC DS-2010-02: **Jonathan Zvesper**
Playing with Information

ILLC DS-2010-03: **Cédric Dégremont**
The Temporal Mind. Observations on the logic of belief change in interactive systems

ILLC DS-2010-04: **Daisuke Ikegami**
Games in Set Theory and Logic

ILLC DS-2010-05: **Jarmo Kontinen**
Coherence and Complexity in Fragments of Dependence Logic

ILLC DS-2010-06: **Yanjing Wang**
Epistemic Modelling and Protocol Dynamics

ILLC DS-2010-07: **Marc Staudacher**
Use theories of meaning between conventions and social norms

ILLC DS-2010-08: **Amélie Gheerbrant**
Fixed-Point Logics on Trees

ILLC DS-2010-09: **Gaëlle Fontaine**
Modal Fixpoint Logic: Some Model Theoretic Questions

ILLC DS-2010-10: **Jacob Vosmaer**
Logic, Algebra and Topology. Investigations into canonical extensions, duality theory and point-free topology.

ILLC DS-2010-11: **Nina Gierasimczuk**
Knowing One's Limits. Logical Analysis of Inductive Inference

ILLC DS-2010-12: **Martin Mose Bentzen**
Stit, Iit, and Deontic Logic for Action Types

ILLC DS-2011-01: **Wouter M. Koolen**
Combining Strategies Efficiently: High-Quality Decisions from Conflicting Advice

ILLC DS-2011-02: **Fernando Raymundo Velazquez-Quesada**
Small steps in dynamics of information

ILLC DS-2011-03: **Marijn Koolen**
The Meaning of Structure: the Value of Link Evidence for Information Retrieval

ILLC DS-2011-04: **Junte Zhang**
System Evaluation of Archival Description and Access

ILLC DS-2011-05: **Lauri Keskinen**
Characterizing All Models in Infinite Cardinalities

ILLC DS-2011-06: **Rianne Kaptein**
Effective Focused Retrieval by Exploiting Query Context and Document Structure

ILLC DS-2011-07: **Jop Briët**
Grothendieck Inequalities, Nonlocal Games and Optimization

ILLC DS-2011-08: **Stefan Minica**
Dynamic Logic of Questions

ILLC DS-2011-09: **Raul Andres Leal**
Modalities Through the Looking Glass: A study on coalgebraic modal logic and their applications

ILLC DS-2011-10: **Lena Kurzen**
Complexity in Interaction

ILLC DS-2011-11: **Gideon Borensztajn**
The neural basis of structure in language

ILLC DS-2012-01: **Federico Sangati**
Decomposing and Regenerating Syntactic Trees

ILLC DS-2012-02: **Markos Mylonakis**
Learning the Latent Structure of Translation

ILLC DS-2012-03: **Edgar José Andrade Lotero**
Models of Language: Towards a practice-based account of information in natural language

ILLC DS-2012-04: **Yurii Khomskii**
Regularity Properties and Definability in the Real Number Continuum: idealized forcing, polarized partitions, Hausdorff gaps and mad families in the projective hierarchy.

ILLC DS-2012-05: **David García Soriano**
Query-Efficient Computation in Property Testing and Learning Theory

ILLC DS-2012-06: **Dimitris Gakis**
Contextual Metaphilosophy - The Case of Wittgenstein

ILLC DS-2012-07: **Pietro Galliani**
The Dynamics of Imperfect Information

ILLC DS-2012-08: **Umberto Grandi**
Binary Aggregation with Integrity Constraints

ILLC DS-2012-09: **Wesley Halcrow Holliday**
Knowing What Follows: Epistemic Closure and Epistemic Logic

ILLC DS-2012-10: **Jeremy Meyers**
Locations, Bodies, and Sets: A model theoretic investigation into nominalistic mereologies

ILLC DS-2012-11: **Floor Sietsma**
Logics of Communication and Knowledge

ILLC DS-2012-12: **Joris Dormans**
Engineering emergence: applied theory for game design

ILLC DS-2013-01: **Simon Pauw**
Size Matters: Grounding Quantifiers in Spatial Perception

ILLC DS-2013-02: **Virginie Fiutek**
Playing with Knowledge and Belief

ILLC DS-2013-03: **Giannicola Scarpa**
Quantum entanglement in non-local games, graph parameters and zero-error information theory

ILLC DS-2014-01: **Machiel Keestra**
Sculpting the Space of Actions. Explaining Human Action by Integrating Intentions and Mechanisms

ILLC DS-2014-02: **Thomas Icard**
The Algorithmic Mind: A Study of Inference in Action

ILLC DS-2014-03: **Harald A. Bastiaanse**
Very, Many, Small, Penguins

ILLC DS-2014-04: **Ben Rodenhäuser**
A Matter of Trust: Dynamic Attitudes in Epistemic Logic

ILLC DS-2015-01: **María Inés Crespo**
Affecting Meaning. Subjectivity and evaluativity in gradable adjectives.

ILLC DS-2015-02: **Mathias Winther Madsen**
The Kid, the Clerk, and the Gambler - Critical Studies in Statistics and Cognitive Science

ILLC DS-2015-03: **Shengyang Zhong**
Orthogonality and Quantum Geometry: Towards a Relational Reconstruction of Quantum Theory

ILLC DS-2015-04: **Sumit Sourabh**
Correspondence and Canonicity in Non-Classical Logic

ILLC DS-2015-05: **Facundo Carreiro**
Fragments of Fixpoint Logics: Automata and Expressiveness

ILLC DS-2016-01: **Ivano A. Ciardelli**
Questions in Logic

ILLC DS-2016-02: **Zoé Christoff**
Dynamic Logics of Networks: Information Flow and the Spread of Opinion

ILLC DS-2016-03: **Fleur Leonie Bouwer**
What do we need to hear a beat? The influence of attention, musical abilities, and accents on the perception of metrical rhythm

ILLC DS-2016-04: **Johannes Marti**
Interpreting Linguistic Behavior with Possible World Models

ILLC DS-2016-05: **Phong Lê**
Learning Vector Representations for Sentences - The Recursive Deep Learning Approach

ILLC DS-2016-06: **Gideon Maillette de Buy Wenniger**
Aligning the Foundations of Hierarchical Statistical Machine Translation

ILLC DS-2016-07: **Andreas van Cranenburgh**
Rich Statistical Parsing and Literary Language

ILLC DS-2016-08: **Florian Speelman**
Position-based Quantum Cryptography and Catalytic Computation

ILLC DS-2016-09: **Teresa Piovesan**
Quantum entanglement: insights via graph parameters and conic optimization

ILLC DS-2016-10: **Paula Henk**
Nonstandard Provability for Peano Arithmetic. A Modal Perspective

ILLC DS-2017-01: **Paolo Galeazzi**
Play Without Regret

ILLC DS-2017-02: **Riccardo Pinosio**
The Logic of Kant's Temporal Continuum

ILLC DS-2017-03: **Matthijs Westera**
Exhaustivity and intonation: a unified theory

ILLC DS-2017-04: **Giovanni Cinà**
Categories for the working modal logician

ILLC DS-2017-05: **Shane Noah Steinert-Threlkeld**
Communication and Computation: New Questions About Compositionality

ILLC DS-2017-06: **Peter Hawke**
The Problem of Epistemic Relevance

ILLC DS-2017-07: **Aybüke Özgün**
Evidence in Epistemic Logic: A Topological Perspective

ILLC DS-2017-08: **Raquel Garrido Alhama**
Computational Modelling of Artificial Language Learning: Retention, Recognition & Recurrence

ILLC DS-2017-09: **Miloš Stanojević**
Permutation Forests for Modeling Word Order in Machine Translation

ILLC DS-2018-01: **Berit Janssen**
Retained or Lost in Transmission? Analyzing and Predicting Stability in Dutch Folk Songs

ILLC DS-2018-02: **Hugo Huurdeman**
Supporting the Complex Dynamics of the Information Seeking Process

ILLC DS-2018-03: **Corina Koolen**
Reading beyond the female: The relationship between perception of author gender and literary quality

ILLC DS-2018-04: **Jelle Bruineberg**
Anticipating Affordances: Intentionality in self-organizing brain-body-environment systems

ILLC DS-2018-05: **Joachim Daiber**
Typologically Robust Statistical Machine Translation: Understanding and Exploiting Differences and Similarities Between Languages in Machine Translation

ILLC DS-2018-06: **Thomas Brochhagen**
Signaling under Uncertainty

ILLC DS-2018-07: **Julian Schlöder**
Assertion and Rejection

ILLC DS-2018-08: **Srinivasan Arunachalam**
Quantum Algorithms and Learning Theory

ILLC DS-2018-09: **Hugo de Holanda Cunha Nobrega**
Games for functions: Baire classes, Weihrauch degrees, transfinite computations, and ranks

ILLC DS-2018-10: **Chenwei Shi**
Reason to Believe

ILLC DS-2018-11: **Malvin Gattinger**
New Directions in Model Checking Dynamic Epistemic Logic